# Exhibit F

# Multilevel Polar Coded-Modulation for Wireless Communications

by

**Hossein Khoshnevis**

A dissertation submitted to the

Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

**Doctor of Philosophy in Electrical and Computer Engineering**

Ottawa-Carleton Institute for Electrical and Computer Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario

September, 2018

# Abstract

In wireless channels, the signal quality degrades mainly due to the additive noise and the random variation of attenuation of the signal, known as fading. The additive noise can be compensated to some extent using forward error correction (FEC) coding and automatic repeat request (ARQ). The fading can be compensated not only with FEC codes and ARQ schemes but also using spatial diversity and multiplexing achieved by employing multiple antenna systems, known as multiple-input multiple-output (MIMO) systems. MIMO schemes fall into different categories based on system requirements, e.g., space-time block codes (STBCs) and limited feedback schemes.

FEC codes, as a method of substantial performance improvement, are employed in most modern communication systems. However, the optimal design of the concatenation of FEC codes and modulation schemes for different applications is an open problem. Polar codes are a new class of FEC codes that benefit from simple rate matching and a variety of low-complexity decoders which facilitate the design of efficient systems. Multilevel coding with multistage decoding (MLC/MSD) is a low-complexity capacity achieving coded-modulation technique that can be designed efficiently for polar codes due to the conceptual similarity.

In this thesis, in order to achieve low-arithmetic-complexity/high-performance coded-modulation schemes for wireless channels, multilevel polar coded-modulation

(MLPCM) schemes are designed and analyzed. To this end, new methods are proposed to decrease the MSD log-likelihood ratio estimation complexity and to improve the set-partitioning based bit-to-symbol mapping (SPM) design. Furthermore, some novel methods are proposed for the low-memory-space/low-arithmetic-complexity design of polar codes in additive white Gaussian noise and slow fading channels. In addition, a set of algorithms are proposed to design MLPCM based on the throughput for hybrid ARQ and channel-adaptive schemes. Moreover, to design the signal for MLC/MSD, a variety of multidimensional constellations are optimized using union bounds on the error rate of uncoded MIMO schemes, and their performance in the presence of polar codes is evaluated. Finally, MLPCM is designed in slow fading channels by optimizing STBCs based on the outage probability, generating SPM based on a bound on the outage, and the joint optimization of polar codes and STBCs.

# Acknowledgments

First and foremost, I want to thank my co-supervisors Professor Halim Yanikomeroglu and Professor Ian Marsland. I am thankful to Professor Yanikomeroglu for the encouragement and the endless support throughout my Ph.D. studies. His advice on both research as well as on my career have been invaluable. I would also like to appreciate Professor Marsland for the enormous contribution of time and ideas to increase the productivity of my thesis. During my studies, we had numerous meetings and discussions on a variety of research issues; throughout them, he has taught me many aspects of communication systems. Completing my Ph.D. would have been difficult without his excellent support and patience.

I would also like to thank my committee members and in particular my external examiner, Prof. Jean-Yves Chouinard, for their invaluable comments and feedback about many aspects of my thesis.

I would like to acknowledge Huawei Technologies Canada Co., Ltd., the Ontario Ministry of Economic Development and Innovations Ontario Research Fund - Research Excellence (ORF-RE) program, and the Natural Sciences and Engineering Research Council of Canada's (NSERC) Strategic Partnership Grants for Projects (SPG-P) program for the financial support of this work. I am grateful to Huawei technologies for filing one patent application from my researches. I especially would like to thank Dr. Ngoc Dao and Dr. Gamini Senarath (Huawei Canada Co., Ltd.) for their invaluable comments and discussions.

I would also appreciate all the members of our group for their encouragement, friendship, and advice. During our numerous research meetings, I learned many lessons from their presentations.

Last but not least, I am thankful to my parents for their encouragement and fusing in me the desire to learn.

# Contents

# List of Tables

# List of Figures

# List of Acronyms

| Acronym | Definition |
|---|---|
| 5G | Fifth Generation Wireless Networks |
| AMC | Adaptive Modulation and Coding |
| ARQ | Automatic Repeat Request |
| AWGN | Additive White Gaussian Noise |
| BEC | Binary Erasure Channel |
| BER | Bit Error Rate |
| BICM | Bit-Interleaved Coded-Modulation |
| BIPCM | Bit-Interleaved Polar Coded-Modulation |
| BLER | Block Error Rate |
| bpcu | Bits-per-Channel-Use |
| BPSK | Binary Phase Shift Keying |
| BSC | Binary Symmetric Channel |
| CPU | Central Processing Unit |
| CRC | Cyclic Redundancy Check |
| CSIT | Channel State Information at the Transmitter |
| CSIR | Channel State Information at the Receiver |
| D-Blast | Diagonal Bell Laboratories Layered Space-Time |

| Acronym | Definition |
|---|---|
| dB | Decibel |
| FEC | Forward Error Correction |
| FER | Frame Error Rate |
| GA | Gaussian Approximation |
| HARQ | Hybrid ARQ |
| IPM | Interior-Point Method |
| IR | Incremental Redundancy |
| KB | Kilobyte |
| KKT | Karush-Kuhn-Tucker |
| LDPC | Low-Density Parity-Check |
| LLR | Log-Likelihood Ratio |
| LMMSE | Linear Minimum Mean Square Error |
| LSD | List Sphere Decoder |
| LTE | Long Term Evolution |
| LTE-A | LTE Advanced |
| MIMO | Multiple-Input Multiple-Output |
| MISO | Multiple-Input Single-Output |
| ML | Maximum Likelihood |
| MLC | Multilevel Coding |
| MLPCM | Multilevel Polar Coded-Modulation |
| MSD | Multistage Decoding |
| NLP | Nonlinear Program |
| OSTBC | Orthogonal Space-Time Block Code |

| Acronym | Definition |
|---|---|
| PAM | Pulse-Amplitude Modulation |
| PEP | Pairwise Error Probability |
| PSO | Particle Swarm Optimization |
| QAM | Quadrature Amplitude Modulation |
| QOSTBC | Quasi-Orthogonal Space-Time Block Code |
| QPSK | Quadrature Phase Shift Keying |
| RAM | Random-Access Memory |
| SBC | Space Block Code |
| SCD | Successive Cancellation Decoder |
| SCLD | Successive Cancellation List Decoder |
| SISO | Single-Input Single-Output |
| SM | Spatial Modulation |
| SNR | Signal-to-Noise Ratio |
| SPM | Set-Partitioning based Bit-to-Symbol Mapping |
| SQPM | Sequential Quadratic Programming Method |
| STBC | Space-Time Block Code |
| STTC | Space-Time Trellis Code |
| SVD | Singular Value Decomposition |
| TCM | Trellis Coded-Modulation |
| TVSBC | Time-varying SBC |
| UBPOP | Upper Bound on the Pairwise Outage Probability |
| V-Blast | Vertical Bell Laboratories Layered Space-Time |

# List of Symbols and Operators

| Symbol | Definition |
|---|---|
| $\mathbb{1}(\cdot)$ | The unit step function |
| $A_R$ | A time-varying rotation matrix |
| $B$ | The number of address bits of a constellation |
| $\mathbf{c}$ | An FEC codeword |
| $c_p(\cdot)$ | The cost of each constellation point in a union bound |
| $c_{p,tot}$ | The total cost of a bit-to-symbol mapping in a union bound |
| $C_\epsilon$ | The total outage capacity for a given $\epsilon$ |
| $C_{\epsilon_i,i}$ | The level-wise outage capacity for a given $\epsilon_i$ |
| $d$ | The value of the an address bit of a constellation |
| $\mathbf{D}_G$ | A pairwise metric matrix |
| $D_H(\cdot)$ | The Hamming distance |
| $E[\cdot]$ | The expectation operator |
| $f_d$ | The normalized Doppler frequency |
| $\mathbf{F}_p$ | The polar code kernel |
| $\mathbf{G}$ | A STBC matrix with elements $g_{n_t}^t$ |
| $\mathbf{G}_X$ | A specific STBC matrix |
| $\mathbf{G}_p$ | The generator matrix of a polar code |

| Symbol | Definition |
|---|---|
| $\mathbf{H}$ | A MIMO flat fading channel matrix with elements $h_{n_t,n_r}$ |
| $\mathcal{H}$ | The Hermitian transpose operator |
| $\mathcal{I}$ | The mutual information |
| $I_0(\cdot)$ | The modified Bessel function of the first kind with zero order |
| $I_{Max}$ | The maximum number of iterations |
| $K_C$ | The total number of two dimensions of a multidimensional constellation |
| $K$ | The message word length |
| $K_0(\cdot)$ | The modified Bessel function of the second kind with zero order |
| $L$ | The number of retransmissions of a HARQ protocol |
| $L_{CRC}$ | The length of a CRC sequence |
| $L_p$ | The list size for a SCLD |
| $\mathbb{L}$ | A large positive number |
| $m$ | The parameter of a Nakagami-$m$ distribution |
| $\mathbf{m}$ | The message word |
| $N$ | The length of a component codeword |
| $N_{tot}$ | The total length of the codeword |
| $N_0$ | The noise power |
| $N_t$ | The number of transmit antennas |
| $N_r$ | The number of receive antennas |
| $N_{SIM}$ | The number of simulated codewords |
| $O(\cdot)$ | The complexity operator |
| $P_{out}$ | The outage probability |

| Symbol | Definition |
|---|---|
| $\mathrm{P}(\cdot)$ | The probability |
| $P_s$ | The symbol error probability |
| $P_b$ | The bit error probability |
| $P_K$ | The FER |
| $q_{i,j}$ | A constant related to UBPOP |
| $Q(\cdot)$ | The Q-function |
| $\mathbf{R}$ | The vector of MLC component code rates |
| $R_i$ | The component FEC code rate |
| $R_0(\cdot)$ | The cutoff rate |
| $R_{tot}$ | The total FEC code rate |
| $\mathbf{S}$ | A specific space-time symbol |
| $s^{(k)}$ | A 2D symbol corresponding to the $k^{th}$ two dimensions of a multidimensional constellation |
| $\mathrm{tr}[\cdot]$ | The trace operator |
| $T$ | The number of time slots of a STBC |
| $v_i$ | The conditional BER of $i^{th}$ bit-channel |
| $\mathcal{T}$ | The transpose operator |
| $\mathbf{W}$ | The additive white Gaussian noise |
| $\mathcal{X}_b^d$ | The set of constellation points with $d = 0$ or $d = 1$ in their $b^{th}$ position given the upper level bits are known for MSD |
| $\bar{\mathcal{X}}_b^d$ | The set of constellation points with $d = 0$ or $d = 1$ in their $b^{th}$ position for BICM |
| $\mathbf{Y}$ | The received signal |

| Symbol | Definition |
|---|---|
| $\mathbf{z}$ | A vector of indices of multidimensional bit-to-symbol mapping |
| $\alpha$ | A parameter of a STBC |
| $\mathcal{A}$ | The information set of a polar code |
| $\beta$ | A parameter of a STBC |
| $\gamma$ | The instantaneous SNR |
| $\bar{\gamma}_{i,j}$ | The average SNR of each path in a MIMO channel |
| $\bar{\gamma}$ | The total average received SNR |
| $\bar{\gamma}_L(\cdot)$ | The lower incomplete Gamma function |
| $\Gamma(\cdot)$ | The Gamma function |
| $\Delta$ | The pairwise distance matrix of a space-time signal |
| $\epsilon$ | The total outage probability |
| $\epsilon_i$ | The outage probability for the $i^{th}$ level of a MLC/MSD scheme |
| $\boldsymbol{\zeta}$ | The vector of parameters of a STBC |
| $\eta$ | The throughput |
| $\bar{\eta}$ | The average throughput |
| $\theta$ | The random rotation angle of a TVSBC |
| $\lambda_b$ | The LLR of the $b^{th}$ address bit of a constellation |
| $\bar{\lambda}$ | The average LLR |
| $\mu$ | A constant related to SNR |
| $\nu$ | The nodes of polar graph |
| $\Pi$ | The multiplication operator |
| $\rho(\cdot), \hat{\rho}_{i,j}$ | A pairwise Bhattacharyya coefficient |
| $\Sigma$ | The summation operator |

# Chapter 1

# Introduction

## 1.1 Motivations

Information transmitted over wireless channels suffers from additive noise and the random variation of the signal attenuation known as fading. The additive noise can be overcome by employing error correction coding and automatic repeat request (ARQ) techniques. The fading, in turn, can be overcome not only by the use of error correction coding and ARQ techniques but also by the use of different fading paths to achieve spatial diversity and multiplexing. To this end, multiple antenna systems, known as multiple-input multiple-output (MIMO) systems, have been employed widely in wireless communication systems. MIMO systems employ signals distributed on antennas and time-slots known as space-time signals.

Forward error correction (FEC) codes are commonly used as one method for significant performance enhancement in modern communication systems including MIMO systems. FEC codes add redundancy to the message allowing the receiver to detect and correct a portion of errors occurred due to the noise or the fading. When the channel is poor (e.g., deep fading), more redundancy is added while when the channel is good, less redundancy should be added to the message word. The amount of the redundancy is controlled in a process known as rate matching. Polar codes, as a new

class of FEC codes, benefit from simple rate matching and a variety of low-complexity decoders. Therefore, they facilitate the design of reliable communication systems. To achieve high performance polar coded schemes, the polar codes, the signal constellations (including space-time constellations), and the method of concatenating the codes and constellations, should be carefully analyzed and designed. In the following, we briefly review some aspects of wireless systems, including coded-modulation schemes and space-time signal design and explain some important open problems in wireless coded-modulation design. In the end, we express the contributions of this thesis to answer some aspects of these problems.

### 1.1.1 Coded-Modulation Techniques

Polar codes, introduced by Erdal Arıkan in [1], are a new class of FECs that work based on the concept of channel polarization to improve some of the equivalent bit-channels at the expense of others. They are provably capacity-achieving codes for binary-input symmetric-output memoryless channels with a successive cancellation decoder (SCD). However, they can practically achieve the capacity of a variety of channels [2]. Polar codes benefit from simple rate matching and low-complexity decoders [3–7] that facilitates the design of effective systems. Therefore, the design of polar coded-modulation may provide us with performance improvements or at least a performance-complexity trade-off in comparison to well-known parallel concatenated (turbo) and low-density parity-check (LDPC) codes.

To achieve the channel capacity with binary codes, generally high order modulations are employed. Therefore, the method of concatenating the channel encoder and modulation is of importance [8]. Among the coded-modulation techniques for binary FEC codes, bit-interleaved coded modulation (BICM) [9] and multilevel coding with multistage decoding (MLC/MSD) [10, 11] have attracted much attention due to their simplicity and good practical performance. BICM consists of one binary

code while MLC/MSD consists of a set of component codes corresponding to different address bits of a constellation. During the last decades, BICM has attracted more attention and has been implemented in most standards due to its perceived simpler implementation, lower decoding delay, and simpler design rules than MLC/MSD systems [12]. For example, while BICM has been included in HSPA [13], LTE-A [14], IEEE 802.11 [15, 16], DVB-T2 [17], DVB-C2 [18] and DVB-S2 [19], MLC has been included only in DRM [20]. However, different methods of coded-modulation design and constellation shaping can result in the superiority of one of these methods in specific conditions [21–23]. For designing polar coded-modulation schemes for the AWGN channel, bit-interleaved polar coded-modulation (BIPCM) and multilevel polar coded modulation (MLPCM) have been evaluated in [24–26], and due to the clarity of design and the conceptual similarity to channel polarization, MLPCM has shown better performance compared to BIPCM. Note that for polar codes, unlike turbo and LDPC codes, the advantages of BICM are no longer valid. Furthermore, the log-likelihood ratio (LLR) for MLPCM can be estimated with lower arithmetic complexity than BIPCM.

When the channel state information at the transmitter (CSIT) is available, employing channel adaptive modulation and coding (AMC) and hybrid automatic repeat request (HARQ) error control protocols can enhance the performance of coded-modulation schemes. HARQ combines the automatic repeat request and FEC coding to provide more-reliable error control schemes. One of the ultimate measures to determine the quality of AMC and HARQ schemes is the throughput defined as the average rate of successful message delivery. Thus, for AMC and HARQ, the FEC code and the modulation should be designed or chosen to maximize throughput. To maximize the throughput by considering the CSIT, rate-matching plays an important role. Due to simple rate-matching without any substantial change in the structure of the decoder, effective polar codes can easily be implemented for AMC and HARQ schemes.

Especially, as shown in this thesis, polar codes can be more efficiently designed based on the throughput than turbo and LDPC codes can. Due to these promising features, in this thesis, a variety of methods for designing more-reliable MLPCM in wireless channels are proposed and evaluated.

### 1.1.2 MIMO Systems and MIMO Coded-Modulation

MIMO systems can substantially enhance the performance in fading wireless channels. As an attractive feature of MIMO systems, increasing the number of transmit or receive antennas can linearly increase the fading channel capacity at high signal-to-noise ratio (SNR) [27, 28]. Therefore, during the last three decades, there has been much research on the design of MIMO schemes, including transmit diversity and multiplexing. The initial works that used transmit diversity are [29, 30]. In these schemes, the signal is transmitted from one antenna in the first time slot and from other antennas in the next time slots. At the receiver, orthogonal differently faded versions of the signal are combined and processed. Since then, a variety of schemes based on the transmit diversity and spatial multiplexing have been proposed. These schemes, depending on the availability of the CSIT, can fall into two main categories: open-loop and closed-loop MIMO schemes.

Open-loop MIMO schemes include a variety of space-time codes and space-time constellations that use fading paths to achieve diversity and multiplexing gains without considering the instantaneous conditions of the channel. Space-time codes are regular structures that transform a set of two-dimensional (2D) symbols into a multidimensional symbol through multiple antennas and time slots. Space-time coding subsumes a variety of schemes including spatial multiplexing, space-time trellis codes (STTCs), space-time block codes (STBCs), and spatial modulation (SM). In spatial multiplexing schemes, parallel substreams of data are transmitted over the fading

channel using different antennas to achieve a multiplexing gain. Diagonal Bell Laboratories layered space-time (D-BLAST) and vertical Bell Laboratories layered space-time (V-BLAST) schemes, introduced by Bell Labs in [31, 32], are the well-known spatial multiplexing schemes that employ SCDs. STTCs are a class of space-time codes used with trellis coding that achieves high performance at the price of high decoding complexity [33]. In contrast, STBCs are designed to achieve full diversity with relatively low decoding complexity. Orthogonal space-time block codes (OSTBCs) are the well-known class of STBCs used in standards, e.g., LTE-Advanced [34], due to their low decoding complexity.

Although, during the last two decades many high-performance classes of STBCs have been proposed (e.g., perfect codes [35, 36]), most of them suffer from high decoding complexity and therefore, they have been used less in specifications. A less-discussed class of STBCs expanded within one time-slot are space block codes (SBCs) that benefit from low decoding complexity. SBCs can achieve error rates close to STBCs especially by employing time-varying rotations during a codeword transmission [37, 38]. However, the optimal design of SBCs has been an open problem.

Closed-loop space-time schemes are instantaneous-channel adaptive. Thus, they can improve the performance and spectral efficiency over their open-loop counterparts, while they also can reduce the decoding complexity. The precoding techniques [39–41], transmit beamforming [42] and antenna selection [43–45] are among the subgroups of the closed-loop MIMO schemes.

MIMO schemes based on the number of antennas are divided to small, moderate, and large MIMO schemes. Small MIMO schemes have typically less than four antennas, moderate MIMO scheme have around ten antennas and large MIMO schemes can employ up to hundreds of antennas. The choice of the scheme depends on the carrier frequency, device size, required data rate, and implementation cost. Small MIMO schemes are employed in applications such as Internet of Things, WiFi, and

Bluetooth. Throughout this thesis, schemes designed for fading channels practically fall into the category of small MIMO schemes. To achieve modest spectral efficiencies, moderate MIMO schemes with eight antennas at the base station are currently used in LTE. Finally, to achieve high spectral efficiencies, large MIMO schemes, e.g., massive MIMO, are proposed for multiuser and millimeter wave line-of-sight scenarios in the fifth generation wireless networks (5G) and beyond.

The concatenation of MIMO schemes and FEC codes, as a method of using both time and space diversity, has been widely studied and a variety of FEC coded MIMO schemes have been proposed (e.g., the concatenation of OSTBCs with FEC codes in [46, 47], and turbo-MIMO in [48]). Among the few notable works on polar coded MIMO schemes, the concatenation of polar and turbo codes with OSTBCs has been compared in [49]; the performance of a polar coded OSTBC scheme with binary phase shift keying (BPSK) modulation equipped with different numbers of receive antennas has been evaluated in [50]; a polar coded SM scheme has been proposed in [51]; and a polar code design method and a bit-to-symbol mapping generation method for irregular multidimensional constellations have been proposed in [52]. Due to the potential of polar codes for practical applications, the design of the optimal concatenation of polar codes and MIMO schemes is an interesting research problem. Despite this fact, investigating a variety of methods and proposing the best method of the coded-modulation design corresponding to each MIMO scheme in a fading channel is one of the important open questions for designing polar coded MIMO systems. Furthermore, the design of open-loop and closed-loop MIMO schemes suitable for MLPCM schemes is of importance.

### 1.1.3 MLPCM Design Elements

When designing MLC/MSD, constellation shaping, bit-to-symbol mapping, and the polar code design method all are of importance. As an important part of the coded-modulation design, a good constellation shape can substantially improve the performance of communication systems. Traditionally, regular one dimensional (1D) pulse-amplitude modulation (PAM) and 2D quadrature amplitude modulation (QAM) constellations have been used in most communication systems including orthogonal and non-orthogonal transmit diversity. However, optimization of the signal constellation shape can provide us with better signaling for specific communication channels and specific system requirements. Space-time constellations can be categorized as either regular or irregular. Regular space-time constellations can be constructed using regular structures such as a space-time code. In contrast, irregular space-time constellations here are defined as multidimensional constellations that cannot be constructed using a regular structure due to irregular shaping.

Typically, due to the fact that MSD achieves the channel capacity, the symbol-wise average mutual information is maximized to design signals for MLC/MSD [53,54] assuming that capacity-achieving FEC codes are employed. However, no closed-form expression has been proposed for the average mutual information in AWGN and fading channels. Therefore, as an alternative, closed-form error bounds or the cutoff rate may be employed to design the signal. In this case, the entire signal can be optimized based on these bounds using contemporary optimization tools. Due to the weakness of these bounds to provide the optimal signal in terms of the mutual information and the difficulty of optimization of signals with large number of variables, signal structures with a low number of parameters, e.g., STBCs, can be optimized directly by maximizing the average mutual information [55]. The average mutual information is a metric for ergodic channels, e.g., AWGN and independent fading channels. For

slow and block fading channels, the outage probability is a bound on the frame error rate (FER) of the system [56]. Therefore, it can be utilized to design the signal in those channels.

For many contemporary applications, short to moderate length codes are employed due to delay requirements and the limitations in decoding complexity. While signal design based on the average mutual information and the outage probability can improve the performance even for non-capacity-achieving FEC codes [57], it does not consider the structure of the FEC code and the decoder. Instead, novel objective functions based on the structure of the FEC codes, decoder, and the bit-to-symbol-mapping may improve the performance of the joint coded-modulation system. Therefore, finding novel methods for designing coded-modulation in the presence of short-to-moderate length FEC codes is an important open problem.

Set-partitioning based bit-to-symbol mapping (SPM) is typically used to enhance the performance of MLC/MSD [58]. SPM can be designed using the Euclidean distance as a metric in AWGN channels [59]. However, for many practical wireless channels including MIMO slow fading channels, the best metric is unknown. Thus, finding an appropriate metric and algorithms for fading channels is of importance since it can substantially enhance the performance of the MLC/MSD scheme.

Due to the simple rate matching of polar codes, they can be adapted to channel conditions. Therefore, for real-time adaptive polar coded-modulation schemes, low-memory-space/low-arithmetic-complexity generation of polar codes is required. Furthermore, system performance is highly dependent on the objective function for the coded-modulation design. Specifically, when CSIT is not available, the average FER can be minimized, and when CSIT is available, AMC can be used and thus, the throughput should be maximized. Therefore, finding a method for low-memory-space/low-arithmetic-complexity design of polar codes based on the desired objective function is an important problem.

## 1.2 Contributions

Throughout this thesis, MLPCM schemes are designed for wireless channels and the performance of MLPCM in terms of the throughput and FER is improved. In addition, the LLR estimation complexity, as the main source of the complexity, is reduced for the MLPCM scheme. Furthermore, SPM and the constellation shape are designed and optimized for various channels and conditions. The contributions of this thesis can be summarized as follows.

- **Simplifying the LLR estimation** – In Chapter 3, a novel method to greatly simplify the LLR estimation of QAM is proposed. The method is based on a simple approach for extending SPM to any size QAM. The proposed method uses a structure similar to polar codes. Furthermore, the LLR estimation for PAM with SPM is simplified.

- **Proposing novel approaches for polar code design** – In Chapter 4, the design of polar coded-modulation using the Gaussian approximation (GA) is simplified and a low-arithmetic-complexity GA-based method is proposed to design the polar coded-modulation for the orthogonal transmission in slow fading channels. Additionally, three low-memory-space/low-arithmetic-complexity methods of the polar coded-modulation design are proposed for the AWGN and the orthogonal transmission in slow fading channels.

- **Throughput-based design of polar codes** – In Chapter 5, polar code and coded-modulation design based on the throughput is proposed to achieve efficient limited length codes for AMC and a variety of HARQ protocols in which CSIT is available. The proposed method is based on low-complexity Gaussian approximation for SCD. Furthermore, since the optimized codes for SCD are not optimal for SC list decoders (SCLD), a rate matching algorithm is proposed

to find the best rate for the SCLD decoders while using the polar codes optimized for SCD. In addition, a novel MIMO MLPCM scheme is proposed based on singular value decomposition (SVD).

- **Multidimensional signal design based on the error bounds** – In Chapter 6, multidimensional space-time signals are designed based on error bounds and compared with STBCs. To this end, a general error bound on the bit error rate (BER) and block error rate (BLER) of uncoded[1] transmission is derived in a Nakagami-$m$ fading channel. In the case of orthogonal transmission, the bound is simplified for SISO and OSTBC and the convexity conditions for 1D constellations are presented. Furthermore, space-time constellations are optimized using the derived bounds and a generalized class of OSTBCs used with different size 2D and multidimensional constellations which provides a variety of performance-complexity trade-offs is proposed. Finally, the pros and cons of employing these constellations to construct MLPCM are studied.

- **Designing multilevel coded-modulation in slow fading channels** – In Chapter 7, STBCs are optimized at low-to-moderate SNRs by minimizing the outage probability for slow fading broadcast channels in which the CSIT is not available. Moreover, a novel class of bounds on the outage probability is derived and used as a criterion to design the SPM in slow fading channels. It is also shown that the MLC/MSD scheme approaches the outage probability of slow fading broadcast channels and for a multistage decoder, the level-wise outage probabilities should be optimally equal under certain conditions. Based on these observations, we derived the outage rule for determining competent code rates for MLC/MSD. In addition, due to the sub-optimality of the available approaches, a novel method is proposed to design the SPM for SBCs in slow

[1]Here, the term uncoded systems is used to point out the systems without any FEC.

fading channels.

- **Proposing a method for the joint optimization of polar codes and STBCs** – In Chapter 7, a method for the joint optimization of MLPCM and STBCs is proposed. In this method, the approximation of the FER achieved by the simulation-based design of polar codes is used as a bound on the performance of the system. Throughout the optimization, the optimizer tries to minimize the average FER of the joint system by choosing a locally optimal match of polar codes, SPM, and parameters of a specific STBC.

## 1.3 Organization

The thesis is organized as follows. In Chapter 2, a generic system model is described. In Chapter 3, algorithms for designing bit-to-symbol mapping and low-complexity LLR estimation for space-time signals including novel methods for PAM and QAM constellations are explained. In Chapter 4, the methods of FER-based polar coded-modulation design including the simulation-based method, the simplified GA-based method for the orthogonal transmission in AWGN and slow fading channels, and novel low-memory-space/low-arithmetic-complexity methods are reviewed and proposed. In Chapter 5, the throughput-based design of polar coded-modulation for AMC and HARQ is proposed. In Chapter 6, multidimensional space-time signals are designed on the basis of the uncoded error bounds and their performance is evaluated in presence of polar codes. In Chapter 7, different aspects of the multilevel coded-modulation design for slow fading channels including STBC design using the outage probability, the bit-to-symbol mapping design using a novel class of bounds on the outage probability, determining the component code rates based on the outage probability, and the joint optimization of polar codes and STBCs are proposed and evaluated. Finally, the conclusions and topics for future work are mentioned in Chapter 8.

# Chapter 2

# Generic System Model

The baseband model of the communication system, as shown in Fig. 2.1, consists of a transmitter and a receiver equipped with $N_t$ transmit and $N_r$ receive antennas, respectively. In the transmitter, each $K$-bit message word $\mathbf{m}$ is encoded using a binary FEC code of rate $R_{tot} = K/N_{tot}$ and length $N_{tot}$, yielding codeword $\mathbf{c}$. Groups of $B$ code bits mapped to a $T \times N_t$ space-time symbol $\mathbf{S}$ with $s_{n_t}^t$ being the matrix element transmitted in time slot $l$ from transmit antenna $n_t$. The space-time symbol is one of the signal points of the space-time modulation signal matrix $\mathcal{S}$ with cardinality $2^B$ and alphabet $\{\mathbf{S}_1, ..., \mathbf{S}_{2^B}\}$. Note that the modulation signal matrix $\mathcal{S}$ can be a space-time code or a space-time constellation. Furthermore, to map the binary code bits to the space-time symbol, a multidimensional bit-to-symbol mapping that spans all symbols of the codebook is employed. The space-time symbol is then sent through the $N_t \times N_r$ time-varying MIMO Rayleigh flat fading channel $\mathbf{H}$ with distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. The $T \times N_r$ received samples $\mathbf{Y}$ at the receiver can be expressed as

$$\mathbf{Y} = \mathbf{SH} + \mathbf{W}, \tag{2.1}$$

where $\mathbf{W}$ is the zero-mean complex additive white Gaussian noise (AWGN) with variance $N_0/2$ per dimension. For fixed links scenarios, the AWGN channel is used

corresponding to $\mathbf{H} = 1$ and $N_t = N_r = 1$. For time-varying scenarios, the slow fading MIMO channel is used in which each codeword $\mathbf{c}$ observes only one fixed uncorrelated realization of $\mathbf{H}$ or a slowly time-varying realization of $\mathbf{H}$ generated using Jakes' method with a normalized Doppler frequency of $f_d$.

The probability of detection for the symbol $\mathbf{S}$, given perfect CSI at the receiver, is proportional to

$$\mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma}) = \frac{1}{\pi^{TN_r}} \exp\left(\frac{-\|\mathbf{Y}\text{-}\mathbf{SH}\|_F^2}{N_0}\right), \tag{2.2}$$

where $\bar{\gamma}$ is the average SNR set to $E[\|\mathbf{S}\|_F^2]/N_0$ and $\|.\|_F$ represents the Frobenius norm. The maximum likelihood (ML) decoding chooses $\mathbf{S}$ such that $\mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma})$ is maximized. For decoding binary codes, instead of symbols, the probability of bits being zero or one is used. Thus, in most decoders, LLR is employed to estimate code bits. The LLR at $b^{th}$ level is defined as

$$\lambda_b = \ln \frac{\mathrm{P}(\mathbf{Y}|c_b = 0, \mathbf{H}, \bar{\gamma})}{\mathrm{P}(\mathbf{Y}|c_b = 1, \mathbf{H}, \bar{\gamma})}, \tag{2.3}$$

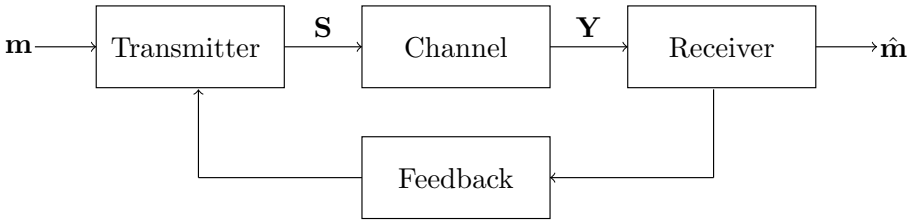



**Figure 2.1:** The system block diagram.

Throughout this thesis, in cases we use the AMC or the HARQ, a feedback stream from the receiver to the transmitter is employed to provide the CSIT or the negative acknowledgment (NACK).

The rest of the chapter is organized as follows: The coded-modulation methods

including MLC/MSD and BICM are described in Sections 2.1 and Section 2.2, respectively. These coded-modulation methods can be used for any class of codes. Due to design and analysis of polar codes in this thesis, the polar encoding and decoding are reviewed in Section 2.3.

## 2.1 Multilevel Encoding/Multistage Decoding

The MLC scheme employs a set of $B$ independent encoders, each with a fixed code length of $N = N_{tot}/B$, with one encoder for each binary channel corresponding to an address bit of a constellation with a cardinality of $2^B$. Each level encoder of the multilevel code encodes a portion of the total $K$ bits corresponding to the component code rates $\{R_1, ..., R_B\}$. After encoding of all levels, each set of code bits $\{c_1^i, c_2^i, ..., c_B^i\}$, for $i = 1, ..., N$, are mapped to a space-time symbol by employing a multidimensional SPM. The total length and the rate of the code are $N_{tot} = BN$ and $R_{tot} = \frac{1}{B}\sum_{b=1}^{B} R_b$, respectively. The structure of MLC/MSD is shown in Fig. 2.2.

In the MSD architecture, code bits of each level are deduced with the aid of the received symbol and previously deduced code bits of the upper levels [10]. As shown in Fig. 2.2, at each level after decoding, the received message word is fed to an encoder and the generated codeword $\hat{\mathbf{c}}^b$ reduces part of the ambiguity for the demapper of the next level to compute more reliable LLRs. Therefore, the probabilities for calculating the LLR in (2.3) are given as

$$\begin{aligned} \mathrm{P}(\mathbf{Y}|c_b = 0, \mathbf{H}, \bar{\gamma}) &= \sum_{\mathbf{S}\epsilon\mathcal{X}_b^0} \mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma}), \\ \mathrm{P}(\mathbf{Y}|c_b = 1, \mathbf{H}, \bar{\gamma}) &= \sum_{\mathbf{S}\epsilon\mathcal{X}_b^1} \mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma}), \end{aligned} \tag{2.4}$$

in which $\mathcal{X}_b^0$ and $\mathcal{X}_b^1$ are the set of all constellation points containing zero or one in

their $b^{th}$ position given $\{\hat{\mathbf{c}}_1, ..., \hat{\mathbf{c}}_{b-1}\}$, respectively.

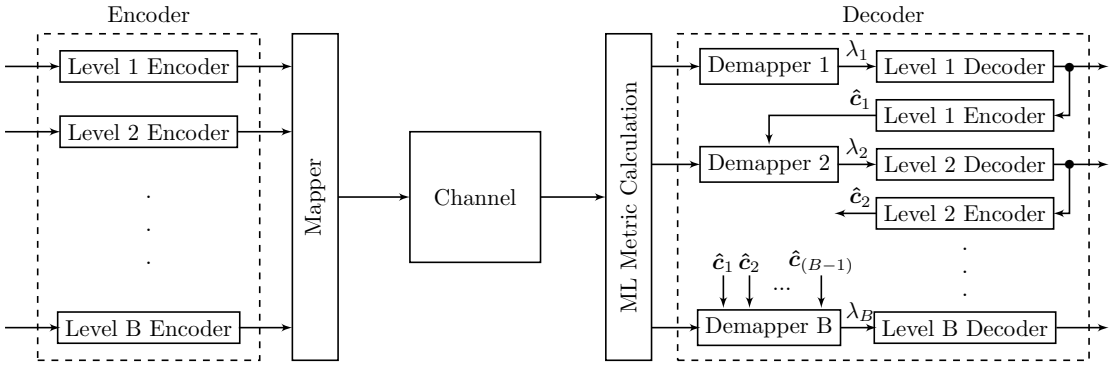



**Figure 2.2:** Multilevel encoding and multistage decoding.

In Fig. 2.2, the ML metric calculation corresponds to estimation of $\mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma})$ for all $2^B$ points. Later, LLRs based on the sets $\mathcal{X}_b^0$ and $\mathcal{X}_b^1$ are generated at demapper of each layer.

## 2.2 Bit-Interleaved Coded-Modulation

In BICM, one binary codeword of length $N_{tot}$ and $R_{tot}$ is interleaved, divided into chunks of $B$ bits, and mapped to a series of space-time symbols using a multidimensional bit-to-symbol mapping. In the decoder, the LLRs of binary channels of a constellation are computed independently. Therefore, the probabilities for calculating the LLR in (2.3) are given as

$$\begin{aligned} \mathrm{P}(\mathbf{Y}|c_b = 0, \mathbf{H}, \bar{\gamma}) &= \sum_{\mathbf{S} \epsilon \bar{\mathcal{X}}_b^0} \mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma}), \\ \mathrm{P}(\mathbf{Y}|c_b = 1, \mathbf{H}, \bar{\gamma}) &= \sum_{\mathbf{S} \epsilon \bar{\mathcal{X}}_b^1} \mathrm{P}(\mathbf{Y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma}), \end{aligned} \tag{2.5}$$

where $\bar{\mathcal{X}}_b^0$ and $\bar{\mathcal{X}}_b^1$ are the set of all space-time symbols with zero or one in their $b^{th}$ position, respectively. In BICM, an additional interleaving is applied on the code-bits prior to bit-to-symbol mapping and the corresponding deinterleaving is applied to the LLRs at the receiver. The BICM block diagram is shown in Fig. 2.3.


**Figure 2.3:** The BICM block diagram.

## 2.3 Polar Codes

In this section, we briefly review the encoder, the SCD and SCLD and the code design method of polar codes. The general discrete-input memoryless channel (DMC) can be written as $\mathrm{W}_C : \mathcal{C} \to \mathcal{Y}$ where $\mathcal{C}$ and $\mathcal{Y}$ denote input and output alphabets, respectively and the probability that $y \in \mathcal{Y}$ is observed given $c \in \mathcal{C}$ is denoted by $\mathrm{W}_C(y \mid c)$. The $N$ independent uses of the channel with probability $\mathrm{W}_C(y_i \mid c_i)$ results in channel $\mathrm{W}_C^N$ that can be expressed as

$$\mathrm{W}_C^N(\mathbf{y} \mid \mathbf{c}) = \prod_{i=1}^{N} \mathrm{W}_C(y_i \mid c_i) \tag{2.6}$$

Arıkan in [1] introduces the polar transformation as a method of linearly transferring $N$ independent uses of the channel to $N$ correlated uses denoted by $\mathrm{W}_C' : \mathcal{U} \to \mathcal{Y}$. The polar transform is applied by the polarization of binary kernel $\mathbf{F}_p$ given by [1]

$$\mathbf{F}_p = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{2.7}$$

By recursively applying the polar transformation on binary DMC input, some of

the resulting correlated code input bits-channels $\mathbf{u}_{1:N} = \{u_1, u_2, ..., u_N\}$[2] are degraded and some are improved. Ideally, for very long codes, the capacity of some bit-channels become one and therefore are used for transmission of information and the capacity of the rest of them become zero and are not used for transmission. When short length codes are used, the capacity of information set bit-channels tend to one and the capacity of bit-channels in complementary set $\mathcal{A}^{\mathcal{C}}$ tend to zero and therefore while long polar codes are practically capacity-achieving, short polar codes perform far from the capacity and have been the subject of much research in last decade to be designed efficiently. Here, for a code of length $N = 2^n$, the first set is called the information set, denoted by $\mathcal{A}$ with cardinality $K$, and the second set is called the frozen set denoted by $\mathcal{A}^{\mathcal{C}}$ with cardinality $N - K$.

### 2.3.1 Encoding

Let $\mathbf{F}_p^{\otimes n}$ be the $n$-th Kronecker power of $\mathbf{F}_p$ defined as $\mathbf{F}_p^{\otimes n} = \mathbf{F}_p \otimes \mathbf{F}_p^{\otimes(n-1)}$. By applying the polar transformation (2.7), the polar codeword $\mathbf{c}$ can be obtained by

$$\mathbf{c} = \mathbf{u}\mathbf{F}_p^{\otimes n} \tag{2.8}$$

Therefore, the generator matrix of the polar code can be defined as $\mathbf{G}_{p,N} = \mathbf{F}_{p,2}^{\otimes n}$. The use of the Kronecker power facilitates constructing the polar encoder since the structure is repeated regularly. As an example, $\mathbf{G}_{p,4}$ is given as

$$\mathbf{G}_{p,4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

[2]Throughout this thesis, since for each expression only a subset of elements of $\mathbf{u}$ are needed, subscript of $\mathbf{u}$ is used to represent the set of indices of available elements.

and $\mathbf{G}_{p,8}$ based on $\mathbf{G}_{p,4}$ can be given as

$$\mathbf{G}_{p,8} = \begin{bmatrix} \mathbf{G}_{p,4} & \mathbf{0}_4 \\ \mathbf{G}_{p,4} & \mathbf{G}_{p,4} \end{bmatrix}$$

where $\mathbf{0_4}$ is a $4 \times 4$ null matrix.

The first part of Fig. 2.4 illustrates how a polar codeword with $N = 8$ by step-wise applying of polar transform $\mathbf{F}_p$ is constructed. Here, the information set and frozen set are $\{u_4, u_6, u_7, u_8\}$ and $\{u_1, u_2, u_3, u_5\}$, respectively.

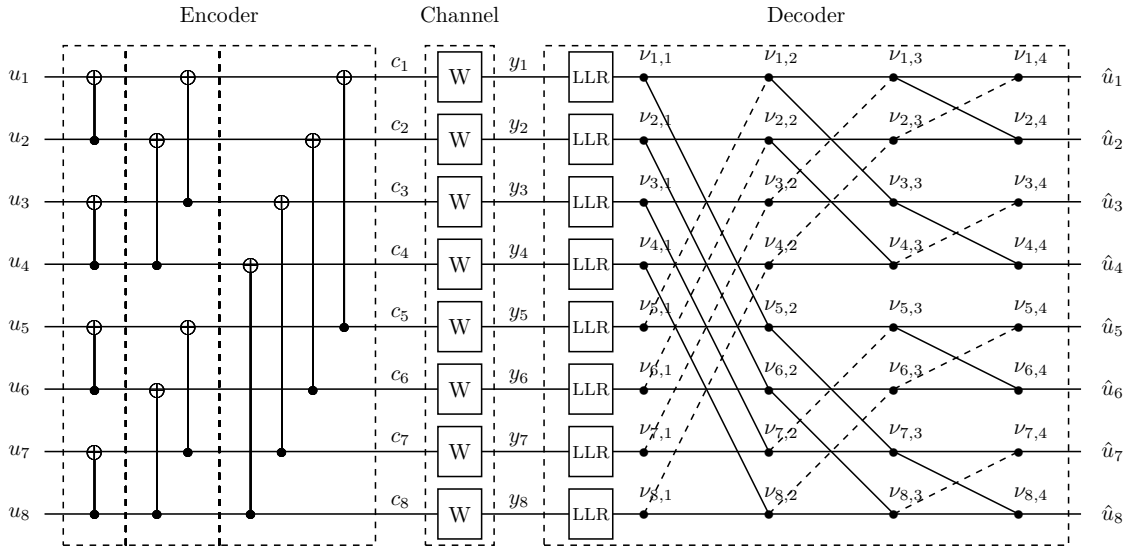



**Figure 2.4:** Polar encoder and decoder.

To encode the polar codes, the kernel (2.7) is applied in $n$ consecutive steps to construct a codeword of length $N = 2^n$ and since in each step, each bit is involved in one operation, the total encoding complexity is of order $O(Nn)$ for implementation of $\mathbf{F}_p^{\otimes n}$ [1]. This low encoding complexity is, in fact, one of the advantages of polar codes in comparison to most of the linear block codes that facilitates the use of these

codes both on hardware and software.

### 2.3.2 Successive Cancellation Decoding

Proposed in [1], SC benefits from low-complexity which is beneficial in many applications. Here, the structure of SC is briefly reviewed. By applying the polar transform, the transition probabilities of the resulting bit-channels, given the code-bit channel output $\mathbf{y}$ and decoded bits $\mathbf{u}_{1:i-1}$ can be computed as [60]

$$\mathrm{W}_{C,i}(\mathbf{y}, \mathbf{u}_{1:i-1} \mid u_i) = \frac{1}{2^{N-1}} \sum_{\mathbf{u}_{i+1:N} \in \{0,1\}^{N-i-1}} \mathrm{W}_C^N\big(\mathbf{y} \mid (\mathbf{u}_{1:i-1}, u_i, \mathbf{u}_{i+1:N})\mathbf{G}_{p,N}\big). \tag{2.9}$$

Denoting the decoder output vector as $\hat{\mathbf{u}}$, the output LLRs given $\mathbf{y}$, $\hat{\mathbf{u}}_{\mathcal{A}^c} = \mathbf{u}_{\mathcal{A}^c}$ and $\mathbf{u}_{1:i-1}$, can be written as

$$\lambda_{i,n+1} = \ln \frac{\mathrm{W}_{C,i}(\mathbf{y}, \hat{\mathbf{u}}_{1:i-1} \mid 0)}{\mathrm{W}_{C,i}(\mathbf{y}, \hat{\mathbf{u}}_{1:i-1} \mid 1)}. \tag{2.10}$$

The final SC decision rule is a hard decision on $\lambda_{i,n+1}$ given as

$$\hat{u}_i = \begin{cases} 0, & \lambda_{i,n+1} \geq 0 \\ 1, & otherwise \end{cases} \tag{2.11}$$

Therefore, the SC decoder (SCD) works based on successive decoding of elements of $\hat{\mathbf{u}}$. The second part of the Fig. 2.4 illustrates the structure of the SCD which is also known as the polar code graph. Each node of this graph is represented by $(i, j)$, where $i \in \{1, .., N\}$ and $j \in \{1, ..., n+1\}$ are row and column indices of the graph. The nodes of graph can be partitioned as type I and type II based on their indices as

$$\mathrm{type}(i,j) = \begin{cases} \mathrm{I} & \lfloor \frac{i-1}{2^{n+1-j}} \rfloor \, (mod\, 2) \equiv 0 \\ \mathrm{II} & \lfloor \frac{i-1}{2^{n+1-j}} \rfloor \, (mod\, 2) \equiv 1 \end{cases}. \tag{2.12}$$

The SCD works based on the soft-hard update decision rules. As shown in Fig. 2.4, the SCD first starts by computing the input LLRs as $\lambda_{i,1} = \ln \frac{W_{C,i}(\mathbf{y} \mid 0)}{W_{C,i}(\mathbf{y} \mid 1)}$. For the intermediate nodes $(2 < j < n+1)$, the LLR update rule of the decoder can be given as

$$\lambda_{i,j} = \begin{cases} \lambda_{i,j-1} \boxplus \lambda_{i^+,j-1} & \text{type I} \\ \lambda_{i,j-1} + (1 - 2\hat{\nu}_{i^-,j})\lambda_{i^-,j-1} & \text{type II} \end{cases}, \tag{2.13}$$

where $i^{\pm} = i \pm 2^{n+1-j}$ and $\hat{\nu}_{i,j}$ is a hard estimate of $\nu_{i,j}$ and the boxplus operator is defined as $\lambda_1 \boxplus \lambda_2 = 2\tanh^{-1}\big(\tanh(\frac{\lambda_1}{2})\ \tanh(\frac{\lambda_2}{2})\big)$ [61]. The LLRs are calculated and passed from left to right through the decoder graph till $\hat{\nu}_{1,n+1}$ is estimated. Then, it starts to pass the hard estimates recursively from right to left to compute the rest of intermediate node LLRs. To estimate $\hat{\nu}_{i,j}$ the polar encoding structure is imitated, i.e., if $\nu_{i,j}$ is type I node, the node is updated by $\hat{\nu}_{i,j} = \hat{\nu}_{i,j+1} \oplus \hat{\nu}_{i^+,j+1}$ and if it is a type II node, $\hat{\nu}_{i,j} = \hat{\nu}_{i,j+1}$.

### 2.3.3 Successive Cancellation List Decoding

For decoding of each output bit with SCD, the information of other previously decoded bits and the future frozen bits are not used. To overcome these shortcomings, a list based decoder has been proposed in [62]. The SCLD records a list containing different possible decoded message words and keeps only $L_p$ most likely paths as it moves ahead from the first to the last message bit. For improving the performance of SCLD, a CRC sequence with the length $L_{CRC}$ is usually added to message bits to increase the probability of finding the most likely message word. It is shown in [62], that by setting $L_p = 32$, the ML performance is achievable at moderate to high SNRs for AWGN channels. Throughout this thesis, for all SCLD design and simulations we use $L_p = 32$.

# Chapter 3

# Bit-to-symbol Mapping Design and LLR Simplification

For matching binary codes to the modulation, an effective coded-modulation scheme should be used. BICM and MLC as two effective coded-modulation schemes benefit from simplicity and good practical performance. A wide range of coded-modulation schemes are indeed variants of these two methods, e.g., [63,64]. BICM and MLC/MSD transfer the communication channel to some parallel binary sub-channels and match binary codes to resulting sub-channels.

MLC/MSD ideally employs codes of different rates for different binary sub-channels of a constellation corresponding to their capacity. MSD is a sequential decoding of adjacent levels of MLC and can provably achieve the channel capacity. The best performance with MLC/MSD is achieved using a bit-to-symbol mapping that increases the variability between bit-channel capacities. A well-known method for this kind of labelling is called the set-partitioning introduced by Ungerboeck for TCM in [59]. Although this algorithm works for regular constellations and small irregular ones, it does not work for large irregular ones. Recently, a new algorithm for designing set-partitioning for irregular multidimensional constellations has been proposed in [52].

BICM employs an interleaver and bit-to-symbol mapping to separate the code and constellation design procedure. Furthermore, it makes an independent selection of the code rate and constellation size possible [9]. However, since the suboptimal LLR estimation, proposed for BICM, increases the information loss, based on the data-processing theorem, BICM suffers from a gap to capacity that can be however to some extent recovered by employing optimal constellation shaping [65] and optimal interleaver design [66]. The best performance with BICM is achieved using a bit-to-symbol mapping that equalizes the capacity of binary sub-channels of a constellation usually called Gray-like mapping.

It has been shown that polar coded-modulation, constructed based on MLC/MSD outperforms polar coded-modulation constructed based on BICM [67] due to the clarity of design and the conceptual similarity of MLC with SPM to channel polarization observed initially in [58]. Moreover, as we will show in this thesis, MLPCM outperforms BICM-based convolutional and turbo coded-modulation schemes as well. Indeed, MLPCM can provide a low-complexity power-efficient scheme that can be employed in a wide range of wireless applications.

In this chapter, we aim to provide algorithms for designing bit-to-symbol mapping for MLC/MSD and BICM and low-complexity methods for the LLR estimation. Thus, for MLC/MSD we start with proposing a simple rule for designing SPM for QAM constellations and based on that, the accurate LLR estimation of QAM with 2D SPM is simplified using a novel device constructed from two independent PAM constellations for the I-channel and Q-channel, and a linear bit mapping. Then, an approximation is proposed to simplify the LLR estimation of PAM constellations. Furthermore, for the general MIMO signaling, we review the low-complexity LLR estimation for MLC/MSD and BICM and the state of the art for SPM generation for MLC/MSD. We further propose an algorithm for designing Gray-like labellings for BICM.

Proposed methods for reducing the LLR estimation complexity of QAM and PAM constellations can be used for the orthogonal transmission in AWGN and fading channels, independent of the number of antennas. These methods can be used for real-time LLR approximation. However, the proposed labelling algorithm for BICM can be used only for small MIMO systems. This method is an offline technique, i.e., it cannot be used for real-time design.

The rest of the chapter is organized as follows: The MLC/MSD structure is explained in Section 3.1. The SPM generation rule and the LLR simplification for QAM and PAM constellations are proposed in Section 3.1.1. The low-complexity LLR estimation and a SPM algorithm for the general form of irregular MIMO signals are reviewed in Sections 3.1.2 and 3.1.3, respectively. The BICM structure is explained in Section 3.2. The low-complexity LLR estimation and a labelling algorithm for irregular MIMO signals are explained in Sections 3.2.1 and 3.2.2, respectively. Finally, conclusions are presented in Section 3.3.

## 3.1 Multilevel Encoding/Multistage Decoding

Imai and Hirakawa in [10], introduced MLC/MSD. The philosophy of MLC/MSD can be explained by using the chain rule of the mutual information. The chain rule of the mutual information for a regular[3] bit-to-symbol mapping is given as

$$\mathcal{I}(\mathbf{Y};\mathbf{c}_1^i,\mathbf{c}_2^i,...,\mathbf{c}_B^i)=\sum_{b=1}^{B}\mathcal{I}(\mathbf{Y};\mathbf{c}_b^i\mid\mathbf{c}_{b-1}^i,...,\mathbf{c}_1^i), \tag{3.1}$$

where $\mathcal{I}(.)$ denotes the mutual information. Obviously, the code-bits of the first level, i.e., $\mathbf{c}_1^i$ are only deduced from $\mathbf{Y}_i$ and then the code-bits of the second level, i.e., $\mathbf{c}_2^i$ are deduced from $\mathbf{Y}_i$ and $\mathbf{c}_1^i$ and this process is continued until the code-bits of the

[3]For the definition of regular mapping refer to [68].

last level, i.e., $\mathbf{c}_B^i$ are deduced.

The structure of MLC/MSD is illustrated in Fig. 2.2. Since the chain rule of the mutual information in (3.1) is sum up to the capacity of the constellation, MLC/MSD can achieve the channel capacity despite its error performance is suboptimal due to the successive decoding of subsequent levels. The comparison of 16-QAM constellation-constraint capacity, and 16-QAM capacity with MLC/MSD and SPM, is shown in Fig. 3.1 for an AWGN channel. It is clear that capacities of binary sub-channels are quite different at moderate SNRs and the total capacity of MLC/MSD with 16-QAM is the same as the 16-QAM constellation-constraint capacity.

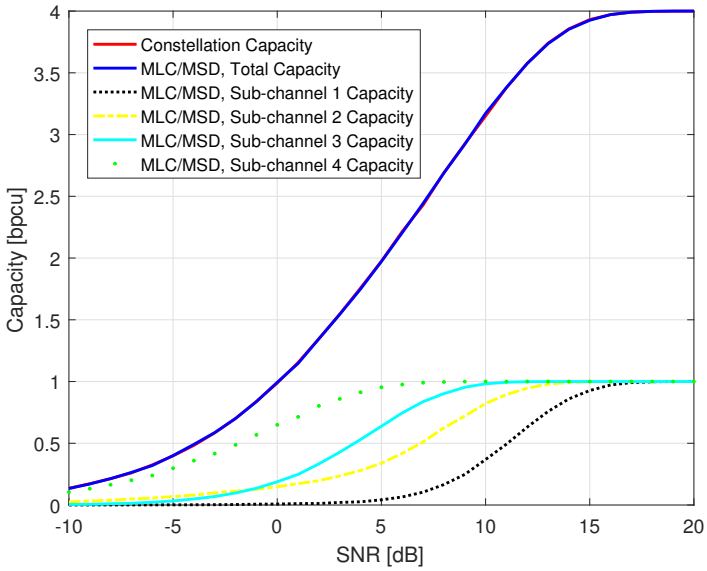



**Figure 3.1:** Comparison of the 16-QAM constellation-constraint capacity and the 16-QAM capacity with MLC/MSD and SPM for an AWGN channel.

In what follows, we derive a set-partitioning rule to design SPM for QAM, and using this rule we simplify the LLR estimation for QAM and PAM. Furthermore, we explain the low-complexity LLR estimation for general irregular multidimensional space-time signals and review the state of the art in designing SPM for multidimensional signals.

### 3.1.1 Low-Complexity LLR Estimation for QAM

Typically, at the decoder of a coded-modulation scheme, calculation of code bit LLRs using (2.3) is much more complex than actually decoding the binary FEC codes. Thus, reducing the LLR calculation complexity can substantially simplify the system. In this section, we simplify the LLR calculation for QAM and PAM constellations with SPM. The results of this section are used throughout the thesis for both decoding and the code construction. Simplified LLR estimation for QAM and PAM can be used for MLC/MSD schemes in AWGN channels or for the orthogonal transmission in fading channels.

#### 3.1.1.1 LLR Calculation for Set-partitioned QAM

The LLR calculation for QAM with a 2D SPM needs a relationship between the LLR value and the real and imaginary parts of the received sample. This results in a function with two input variables and high arithmetic complexity. While this function can be divided into several regions and piecewise approximations with planes can be employed to simplify the LLR calculation, the number of regions grows fast for large size QAM constellations. As an alternative, a 2D SPM can be decomposed into two independent 1D SPMs. While the decomposition of 2D to 1D mappings is well-known and straightforward for Gray mapping, in the following we proposed a new technique that is suitable for SPM.

We want to map $B$ bits, $[c_1 c_2 ... c_B]$ onto a $2^B$-point square QAM constellation with SPM, such that bit $c_1$ has the lowest reliability, followed by $c_2$, and so on up to $c_B$ which is the most reliable. This can be achieved using the simple device shown in Fig. 3.2, that involves a simple code and two PAM symbol mappers with natural mapping. By using this device at the transmitter, the LLR calculator at the receiver is greatly simplified. Using this device, we first precode the bits, giving $[b_1, b_2, ..., b_B]$

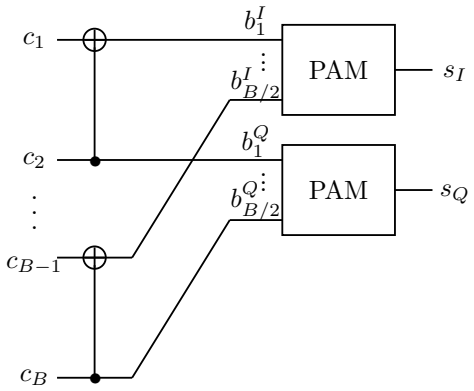



**Figure 3.2:** A transform to map natural numbers to 16-QAM with 2D SPM using two independent 4-PAMs with SPM.

where

$$b_k = \begin{cases} c_k \oplus c_{k+1} & \forall \text{ odd } k, \\ c_k & \forall \text{ even } k. \end{cases} \tag{3.2}$$

Then we send the bits in odd-indexed positions to a $2^{B/2}$-PAM symbol mapper for transmission over the I-channel, and the other bits to another $2^{B/2}$-PAM symbol mapper for the Q-channel. From Fig. 3.2, $b_k^I = b_{2k-1} = c_{2k-1} \oplus c_{2k}$ and $b_k^Q = b_{2k} = c_{2k}$.

The PAM symbol mappers are identical and employ natural mapping between the input bits $\mathbf{d} = [d_{B/2}...d_2 d_1]$ (where $d_k = b_k^I$ or $b_k^Q$ for the I-channel and Q-channel mappers, respectively), and the points

$$\text{SM}_{\text{PAM}}[\mathbf{d}] = 2d - (2^{B/2} - 1), \tag{3.3}$$

where $d = d_{B/2}2^{B/2-1} + ... + d_2 2^1 + d_1 2^0$ is the integer representation of $\mathbf{d}$. Equivalently, we can write

$$\text{SM}_{\text{PAM}}[\mathbf{d}] = \sum_{k=1}^{B/2} (2d_k - 1)2^{k-1}. \tag{3.4}$$

Note that with this mapping, the least significant bit, $d_1$, is also the least reliable,

and the most significant bit, $d_{B/2}$, is the most reliable. The transmitted QAM symbol is then

$$\begin{aligned}\mathrm{SM}_{\mathrm{QAM}}[\mathbf{c}] =& \mathrm{SM}_{\mathrm{PAM}}[\mathbf{b}^I] + \jmath \mathrm{SM}_{\mathrm{PAM}}[\mathbf{b}^Q] \\ &= \sum_{k=1}^{B/2} (2b_k^I - 1)2^{k-1} + \jmath \sum_{k=1}^{B/2} (2b_k^Q - 1)2^{k-1} \\ &= \sum_{k=1}^{B/2} (2[c_{2k-1} \oplus c_{2k}] - 1)2^{k-1} + \jmath \sum_{k=1}^{B/2} (2c_k - 1)2^{k-1},\end{aligned} \tag{3.5}$$

where $\jmath = \sqrt{-1}$.

**Theorem 1.** *The resulting constellation in (3.5) uses SPM.*

*Proof.* Observe that

$$\begin{aligned}\mathrm{SM}_{2^B\text{-QAM}}[\mathbf{c}] = & \sum_{k=1}^{B/2-1} (2[c_{2k-1} \oplus c_{2k}] - 1)2^{k-1} + \jmath \sum_{k=1}^{B/2-1} (2c_k - 1)2^{k-1} \\ & + (2[c_{B-1} \oplus c_B] - 1)2^{B/2-1} + \jmath(2c_B - 1)2^{B/2-1} \\ & = \mathrm{SM}_{2^{B-2}\text{-QAM}}[\mathbf{c}_{1:B-2}] + \mathrm{SM}_{4\text{-QAM}}[\mathbf{c}_{B-1:B}]2^{B/2-1},\end{aligned} \tag{3.6}$$

so that the $2^B$-QAM constellation can be constructed by enlarging a $2^{B-2}$-QAM constellation by a factor of 4. Under the assumption that the $2^{B-2}$-QAM constellation has SPM, then the enlarged constellation does as well. Each point $s = \mathrm{SM}_{2^{B-2}\text{-}QAM}[\mathbf{c}_{1:B-2}]$ in each subset of the smaller constellation is replaced by four points in the larger constellation, $s+(-1-\jmath)2^{B/2-1}$, $s+(1+\jmath)2^{B/2-1}$, $s+(1-\jmath)2^{B/2-1}$, $s+(-1+\jmath)2^{B/2-1}$, with the same bit labellings as $s$ (i.e., $\mathbf{c}_{1:B-2}$), but with 00, 01, 10 or 11 appended, respectively. Since all the points in each set-partitioned subset of the smaller constellation fall on some regular lattice, and the new points in the enlarged subset fall on the same lattice, the minimum distance between points in the subset remains the same. By adding two more layers of the set-partitioning, creating subsets with minimum distances of $2^{B/2}$ and $\sqrt{2}\times 2^{B/2}$, the resulting $2^B$-QAM constellation has SPM, provided the $2^{B-2}$-QAM

constellation has SPM. But since by inspection, the 4-QAM constellation given by (3.5) with $B = 2$ has SPM, it follows by induction that the $2^B$-QAM constellation has SPM for all even values of $B$. ■

Note that the constellation with SPM in (3.5) is regular, since all subsets at a specific level $n$ have the same average Euclidean distance spectrum and consecutively the same capacity [69]. Therefore, one binary code can be constructed for all subsets within each level.

At the receiver, the LLRs of the code bits, $\mathbf{c}$, are readily computed from the received sample. The real and imaginary parts of the received sample are sent to different $2^{B/2}$-PAM LLR calculators, which can use (2.3) to calculate the LLRs corresponding to $\mathbf{b}^I$ and $\mathbf{b}^Q$, respectively. The LLRs of the first PAM bit-channels are calculated, giving $\lambda_1^I$ and $\lambda_1^Q$, which are combined to give the LLR of $c_1$,

$$\lambda_1^c = \lambda_1^I \boxplus \lambda_1^Q. \tag{3.7}$$

Once the upper code (of which $c_1$ is a code-bit) has been decoded, and an estimate $\hat{c}_1$ of $c_1$ based on the code has been generated, the LLR of $c_2$ can be calculated as

$$\lambda_2^c = (1 - 2\hat{c}_1)\lambda_1^I + \lambda_1^Q, \tag{3.8}$$

and the second code can be decoded, giving $\hat{c}_2$. Armed with $\hat{c}_1$ and $\hat{c}_2$, the LLR calculators can detect the second PAM bit-channels, giving $\lambda_2^I$ and $\lambda_2^Q$. These, in turn, are used to calculate

$$\begin{aligned} \lambda_3^c &= \lambda_2^I \boxplus \lambda_2^Q, \\ \lambda_4^c &= (1 - 2\hat{c}_3)\lambda_2^I + \lambda_2^Q. \end{aligned} \tag{3.9}$$

This process is repeated until all levels have been decoded.

#### 3.1.1.2 LLR Calculation for Set-partitioned PAMs

In the last section, we showed how to separate a QAM constellation with 2D SPM into two independent PAM constellations, which by itself significantly reduces the complexity of LLR calculations, but the complexity is still needlessly high. Although the LLR calculation for BPSK is easy, computed as $\lambda = -\frac{4y}{N_0}$ for the AWGN channel, for PAM constellations direct application of (2.3) is of high complexity even for moderately sized constellations. In this section, we try to simplify the LLR estimation of PAM constellations.

The LLR calculation for MSD based on the Jacobi theta functions is proposed in [70], in which knowledge of the values of Jacobi theta functions is required for the LLR calculation. However, saving and referring to values of these functions needs large memory capacity. Instead, the LLR of MSD can be approximated by the dominant term since $\ln(\sum_t e^{-|x_t|}) \approx -\min_t(|x_t|)$ [71]. Therefore, the LLR of MSD in (2.3) can be approximated by

$$\lambda_b \approx \frac{1}{N_0}\left[-\min_{s\in\mathcal{X}_b^0}|y-s|^2 + \min_{s\in\mathcal{X}_b^1}|y-s|^2\right]. \tag{3.10}$$

This approximation, known as the max-log approximation (MLA) has been used for LLR estimation of Gray-mapped QAM [72]. In [73], three methods, namely an MLA in (3.10), a log-separation algorithm (LSA) and a mixed algorithm, are proposed to reduce the complexity of the LLR calculation of MSD for phase shift keying (PSK) and amplitude-phase-shift keying (APSK) constellations. In [73], it is shown that LSA is slightly better for low SNRs, for moderate-to-high SNRs the MLA is essentially the same as the exact calculation using (2.3). Further analysis of the MLA can result in substantial simplification of the LLR calculation. Here, by analyzing the MLA and employing the piecewise linear approximation of the LLRs, we reduce the LLR calculation complexity for PAM constellation with SPM.

**Theorem 2.** *Let $\mathcal{X}_b$ be a $M_b = 2^{B-b+1}$-point subset of a $2^B$-PAM constellation given the $b-1$ upper-level (least significant) bits are known, and let $s_d$ be the $(d+1)^{th}$ largest element of $\mathcal{X}_b$. For $d \in 0, 1, ..., M_b - 2$ and $\mathbb{L}$ a large number, let $\Omega_d = \{y \in \mathbb{R} \mid -\mathbb{L}\delta_d + s_d < y \leq s_{d+1} + \mathbb{L}\delta_{d-M_b+2}\}$[4], be the interval $(s_d, s_{d+1}]$ except the first interval extends down to $-\infty$ and the last up to $+\infty$. Then, for $y \in \Omega_d$*

$$\lambda_b(y) \approx \frac{1}{N_0}(-1)^d(s_{d+1} - s_d)(s_{d+1} + s_d - 2y). \tag{3.11}$$

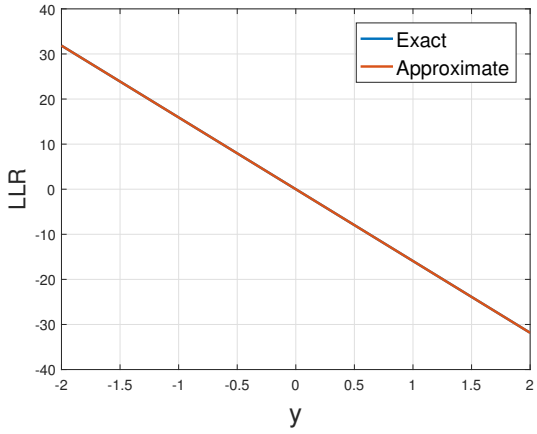
*Proof.* Since $\mathcal{X}_b$ is just a $M_b$-PAM constellation with natural mapping that has been scaled by $2^{b-1}$ and shifted by an amount that depends on the known upper level bits, the label of the least significant bit changes with each consecutive point. Therefore, for $y \in \Omega_d$ the pair $(\min_{s\in\mathcal{X}_b^0}|y-s|^2, \min_{s\in\mathcal{X}_b^1}|y-s|^2)$ is either $(s_d, s_{d+1})$ or $(s_{d+1}, s_d)$ with the first applying when $d$ is even and the second when $d$ is odd. Thus (3.10) can be written as $\frac{1}{N_0}(-1)^d[-|y-s_d|^2 + |y-s_{d+1}|^2]$, which reduces to (3.11). ∎

The LLR as a function of the received symbol $y$, approximated by (3.11) is compared with the exact LLR computed by (2.3) as shown in Fig. 3.3, for the first level of 4-PAM, 8-PAM and BPSK with $\gamma = 6$ dB. This approximation is quite accurate at SNRs of interest but requires much less effort to evaluate than the exact expression.

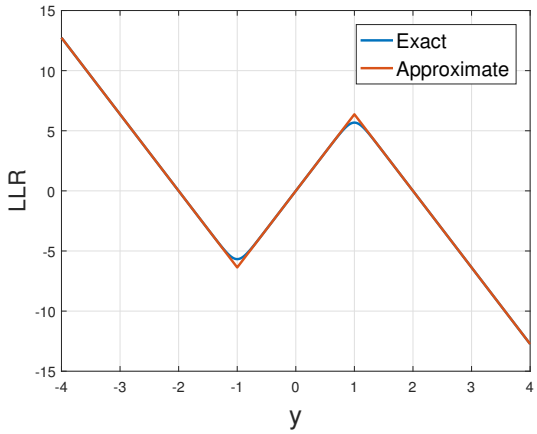
### 3.1.2 Low-Complexity LLR Estimation for Space-Time Signals

The methods proposed in Section 3.1.1 can be used for low-complexity LLR estimation of orthogonal transmission schemes in fading channels. For non-orthogonal schemes, the max-log approximation can be used to simplify the LLR estimation for multilevel space-time signals. This approximation is quite accurate at moderate-to-high SNRs.

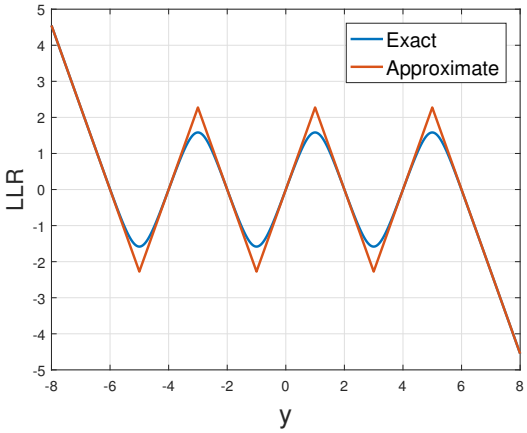
[4]Here, $\delta_d$ represents the Dirac delta function.




(a)




(b)




(c)

**Figure 3.3:** Comparison of the exact and the approximate estimation of LLRs as a function of $y$ for a) a BPSK, b) the first level of 4-PAM, and c) the first level of 8-PAM.

In this case, the max-log approximation for MIMO channel can be given as

$$\lambda_b \approx \frac{1}{N_0}[-\min_{s\epsilon\mathcal{X}_b^0}\|\mathbf{Y\text{-}SH}\|_F^2 + \min_{s\epsilon\mathcal{X}_b^1}\|\mathbf{Y\text{-}SH}\|_F^2]. \tag{3.12}$$

Note that for searching space-time points, low-complexity sphere decoding techniques can be employed. Two well-known sphere decoders are list sphere decoder (LSD) proposed in [74] and soft-to-hard sphere decoder proposed in [75]. The LSD generates a list that makes $\|\mathbf{Y\text{-}SH}\|_F^2$ smallest and searches the list to estimate the LLR. In MSD, the LSD is employed in a level-wise manner, i.e., it only searches within the zero and one symbol sets in each level given the upper-level bits are known. Thus, only the first level search has the complexity of the LSD for BICM. As another well-known sphere decoder, the soft-to-hard decoder approximates the LLR based on the knowledge achieved through hard sphere decoding of the signal. Reducing the complexity of these techniques by limiting the sphere radius or simplifying the approximations results in a penalty in terms of the system performance.

### 3.1.3 Labelling Algorithm for Space-Time Signals

To achieve the best performance with MLC/MSD, the variability between sub-channel capacities should be increased by employing labelling algorithms. In MLC, a bit-to-symbol mapping is needed that matches the level-code rates to the code-bit channel capacities, e.g., assigns the lowest rate code to the lowest capacity code-bit channel and continue subsequentially until assigning the highest rate codes to the highest capacity code-bit channel. A well-known method for this kind of labelling is called set-partitioning introduced by Ungerboeck for trellis-coded modulation (TCM) in [59]. In order to differentiate between bit-channel capacities, Ungerboeck tried to partition a set of length $2^B$ in consecutive steps $b = 1, 2, ..., B$ to subsets of length $2^{B-b+1}$ that at each step the minimum of Euclidean distance between each two point in

the subset is greater than the minimum Euclidean distance in the mother-subset. Ungerboeck's algorithm indeed tries to provide $B$ bit-channels with quite different capacities. Despite this algorithm works for regular constellation and small irregular ones, it does not work for large size irregular ones [52]. Forney in [76] introduced the coset codes that subsumes the trellis code including TCM and lattice codes that are block codes and can be constructed by using MLC. Therefore, he provided a formal set-partitioning algorithm that however only is suitable for regular multidimensional constellations that lie on a lattice. Furthermore, an algorithm for finding labelling for a specific polar code using the exhaustive search of effective permutations has been proposed in [77]. However, searching a large number of possible labellings for medium to large size constellation is infeasible.

Recently, for general irregular multidimensional constellations, in [52] a new set-partitioning algorithm is proposed that works as follows. At the first step, for each symbol, the most distant symbol is found and the minimum of them is considered as a distance threshold for pairing the sets denoted as $\tau$. Then every symbol is paired with the closest another symbol that has the minimum distance of $\tau$ and labels 0 and 1 are assigned to the first and the second symbols in each pair. The threshold $\tau$ ensures that the set of distances of paired points does not have a high variance. After pairing every two symbols, the distance of each two-symbols sets with respect to other two-symbols sets is measured and the minimum of the maximum of their distance is chosen as the new $\tau$ and the mentioned procedure is repeated in $B$ consecutive steps. In [52], the Euclidean distance has been used as a measure of distance between subsets.

## 3.2 Bit-Interleaved Coded-Modulation

BICM is one of the widely used coded-modulation techniques that, unlike MLC, can separate the design of code and constellation [9]. BICM typically employs a bit-to-symbol mapping to equalize the protection of binary sub-channels of the constellation as much as possible and an interleaver to adapt the code-bit protections to binary sub-channels of the constellation. Despite the information loss in typical BICM schemes [9], it can provide good performance in most of the systems. BICM design procedure has two steps: designing a bit-to-symbol mapping and designing an interleaver.

For BICM, a bit-to-symbol mapping should be designed for the constellation to equalize the capacity of different parallel binary sub-channels of a constellation. For regular QAM constellations, Gray mapping and for irregular constellations Gray-like mappings are typically used.

BICM also employs an interleaver that can assign different code-bits to binary sub-channels of constellations by considering their levels of protection. Typically, a single interleaver is utilized in BICM. However, a series of interleavers can be employed to provide a better match between code-bits and the level of protection of parallel binary sub-channels of a constellation [78]. To find the optimal interleaver, typically a bound on the performance of the coded-modulation scheme is used, e.g., such bounds for convolutional or turbo code are employed in [78, 79]. However, the corresponding bounds do not exist for finite length polar codes, and this makes the design of interleaver quite difficult. In this situation, to separate the coding and the modulation, one of the options for designing BICM is to use a random interleaver in the price of the performance loss [80].

To reduce the BICM gap to the capacity for long codes, the interleaver and the bit-to-symbol mapping can be designed for optimal capacity-achieving constellations [53]. The comparison of the 16-QAM constellation-constraint capacity and 16-QAM

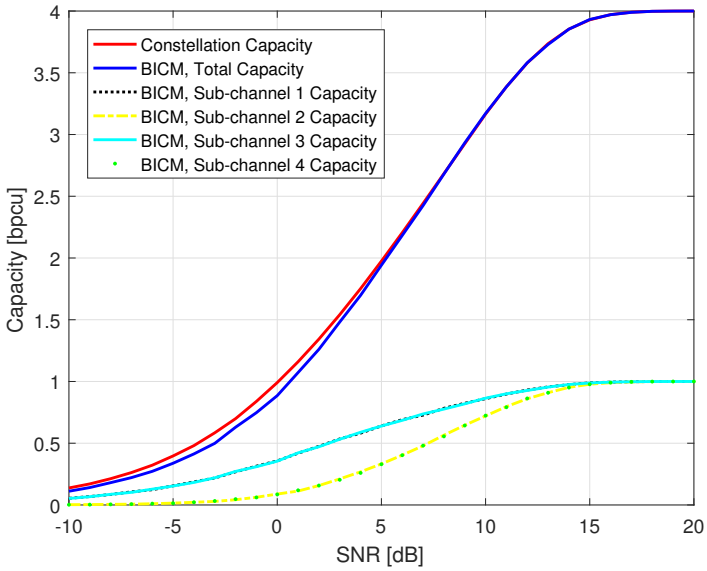



**Figure 3.4:** Comparison of a 16-QAM constellation total capacity and the BICM capacity with a Gray mapped 16-QAM constellation for an AWGN channel.

capacity with BICM and Gray mapping is provided in Fig. 3.4 for an AWGN channel. It is clear that there is a gap between the total BICM capacity and the 16-QAM constellation-constraint capacity.

### 3.2.1 Low-Complexity LLR Estimation for Space-Time Signals

The max-log approximation can be used to simplify the LLR estimation for MIMO BICM schemes [81]. The max-log approximation for BICM can be given as

$$\lambda_b \approx \frac{1}{N_0}\Big[-\min_{s\in\bar{\mathcal{X}}_b^0}\|\mathbf{Y\text{-}SH}\|_F^2+\min_{s\in\bar{\mathcal{X}}_b^1}\|\mathbf{Y\text{-}SH}\|_F^2\Big]. \tag{3.13}$$

In [82], for BICM schemes, it is shown that in price of degradation of the LLR

accuracy, the complexity of the LLR estimation can be further reduced by approximating the LLR using appropriate functions.

### 3.2.2 Labelling Algorithm for Space-Time Signals

In coherent systems, the multidimensional Cartesian product of the Gray-mapped PAM constellation can construct a Gray mapping for the Cartesian product of PAM constellations, known as cubic constellations [83], when used in orthogonal transmission. However, for other classes of regular and irregular constellations, the mapping can be designed based on the position of the points and the channel statistics using a bound on the BER performance of the system.

Due to the relationship between the BER and the LLR estimation for BICM, a good bit-to-symbol mapping can be designed by minimizing the union bound on the BER. By applying the binary switching algorithm in [84–86] to minimize the BER, the best bit-to-symbol mapping can be found. To find the mapping, we modify the binary switching algorithm to adapt it to our problem. The labelling algorithm can be described as Algorithm 1, where $c_p(v)$ is the cost of each symbol equal to the pairwise BER and the total cost can be calculated as the sum of the cost function of each symbol and can be expressed as

$$c_{p,tot} = \sum_{v=1}^{2^B} c_p(v). \tag{3.14}$$

The binary switching algorithm is initialized by a random index vector. The individual cost of points for each index $v$ in constellation and the total cost of the bit-to-symbol mapping is computed using the *UPDATE_COST()* function based on the pairwise error probability (PEP) and (3.14), respectively. The individual pairwise cost of points for each point $v$ are first sorted out using the *SORT_INDICES()* function in decreasing order. Then, the algorithm swaps the index of the point with the highest

individual cost with all other points using *SWITCH_INDEX()* to find a mapping with a lower total cost. Indeed, it is assumed that the highest individual cost should be suppressed first. In case no improvement is achieved by switching the indices, the same procedure is repeated for the rest of the sorted indices in $\mathbf{z}$ in decreasing order. If a better bit-to-symbol mapping is found, the Algorithm 1 starts the next iteration. In case of no improvement after checking all indices in $\mathbf{z}$, the algorithm is halted. In this algorithm, $\mathbf{F}$ is the vector of sorted costs of constellation points, $\mathbf{z}$ and $\mathbf{z}'$ are vectors of indices of constellation points, $F_{tot}$ is the total cost value, $F_{tot,Min}$ is the minimum of the total cost values, and $I_{Max}$ is the maximum number of iterations.

By switching the indices, any index vector can be achieved from any other index vector [84]. Therefore, the globally optimal solution is not out of the achievable range of the solutions, although it is hard to achieve. Here, we choose to start the new iteration after finding an improved bit-to-symbol mapping instead of checking all elements of $\mathbf{z}$ to avoid the greediness. We also examined the case of choosing the best mapping by checking all sorted elements of $\mathbf{z}$ in each iteration which resulted in worse mappings for many constellations due to being more greedy.

To achieve a better result in limited time, Algorithm 1 can be excited many times with different initial points specified by different random vectors, $\mathbf{z}$. Here, we choose to have 10000 different initial points since more than that rarely improves the quality of optimization for small to medium-size constellations.

## 3.3 Conclusion

In this chapter, we simplified the LLR estimation for $2^B$-QAM constellations with SPM using a new device constructed from two independent $2^{B-1}$-PAM constellations and simplified the LLR estimation for PAM. The method can also be used to construct

**Algorithm 1:** Binary Switching Algorithm

**Input :** SER optimized constellation points $\mathbf{s}$

**Output:** A locally optimum bit-to-symbol mapping $\mathbf{z}^*$ for $\mathbf{s}$

**Procedures**:

SORT_INDICES(): Sorts out and returns a vector of indices of constellation points in decreasing order of their cost.

SWITCH_INDEX(i,j): Switch the index of $i^{\text{th}}$ and $j^{\text{th}}$ constellation points in $\mathbf{s}$ and returns the new index vector.

UPDATE_COST(): Calculates and returns the cost of each point based on PEP and the total cost based on (3.14).

**Initialisation**:

Randomly choose an index vector $\mathbf{z}$ for constellation points $\mathbf{s}$ and sort them out based on the random index vector.

$\mathbf{z}^*$=$\mathbf{z}$

**for** *iter* = 1 *to* $I_{Max}$ **do**

  [$F_{tot,Min}$,$\mathbf{F}$]=UPDATE_COST($\mathbf{s}(\mathbf{z}^*)$)

  $\mathbf{z}$=SORT_INDICES($\mathbf{F}$,$\mathbf{z}^*$)

  v=1

  Indicator=0

  IterationFinishFlag=0

  **while** *IterationFinishFlag==0* **do**

    **for** $v'$ = 1 *to* $2^B$ **do**

      **if** *(*$\boldsymbol{z}(v) \neq v'$*)* **then**

        $\mathbf{z}'$=SWITCH_INDEX($\mathbf{z}(v)$,$v'$)

        [$F_{tot}$,.]=UPDATE_COST($\mathbf{s}(\mathbf{z}')$)

        **if** *(*$F_{tot} < F_{tot,Min}$*)* **then**

          $F_{tot,Min} = F_{tot}$

          $\mathbf{z}^* = \mathbf{z}'$

          Indicator=1

      SWITCH_INDEX($v'$,$\mathbf{z}(v)$)

    **if** *(Indicator==1)* $\|$ $(v == 2^B)$ **then**

      IterationFinishFlag=1

    **else if** *Indicator==0* **then**

      v=v+1

return $\mathbf{z}^*$

SPM for QAM with any cardinality. We explained the LLR estimation for MLPCM and BICM used with space-time signals and reviewed the state of the art algorithm for designing SPM. In addition, we proposed a modified binary switching algorithm to design Gray-like mapping for BICM.

# Chapter 4

# Frame Error Rate Based Design for Polar Coded-Modulation

Polar codes are a provably capacity-achieving class of codes for binary-input symmetric-output memoryless channels with the SCD [1]. This includes a wide range of practical channels such as binary symmetric channel (BSC), binary erasure channel (BEC) and binary-input additive white Gaussian noise. Although Arıkan's polar codes are not universal under successive cancellation decoding[5] [88], they still show very good performance in a variety of channels. The simple repetitive structure of polar codes facilitates the encoding and decoding processes, the code design and the rate selection. These promising features make them a good candidate for future communication systems [89]. To achieve the symmetric capacity of the channel with SC decoding, polar codes of very long frame[6]-lengths should be designed. However, in most practical scenarios, we need short-to-moderate frame-lengths to meet the complexity and the delay requirements.

To design the polar code the information set should be determined. Arıkan in [1]

[5]However, polar codes are still good codes under ML decoding, since codes with a good performance under BSC are also good for any channel with the same capacity and in this sense capacity-achieving polar codes for BSC are practically universal [2]. However, there is no low-complexity ML decoder for polar codes over arbitrary channels. Recently, in [2, 87], three different methods of constructing universal modified polar codes are presented.

[6]Throughput this thesis, each modulated FEC codeword is called a frame.

proposed to use a bound based on the Bhattacharryaa parameter for designing polar codes in BECs that can be computed recursively. He also mentioned the Monte Carlo based simulation of polar code as a method of determining the information set for an arbitrary channel. However, the efficient design of polar codes for an arbitrary channel has been an active research field. Due to the fixed polar graph, Mori and Tanaka in [90] proposed to use the density evolution to estimate the information set for a SCD. Later, Kern et al., in [91], tried to simplify the density evolution using the min-sum approximation of the boxplus operation that results in performance degradation compared to the density evolution and still needs a very large space.

Due to the unreasonable complexity of density evolution, Tal and Vardy in [60] proposed a low-complexity finite output alphabet design of polar codes. However, for the extension to the AWGN channel, we should resort to quantization with a large number of levels which in turn results in high complexity. Trifonov in [92] proposed the Gaussian approximation (GA) of the distribution of LLRs as a low-complexity version of the density evolution for AWGN channels. Although, as shown in [91, 93] the performance of GA is approximately the same as the perfect density evolution, it has been only used to design the code in an AWGN channel. In [94], the simulation-based design of polar codes for the binary-input Rayleigh channel has been discussed when the CSIR and CSIT and channel statistics are available. In [95], a code design method is proposed based on an approximation of bit-channel error probabilities for fast fading channels, when the channel statistics and CSIR are available.

Designing of FEC coded schemes for multidimensional constellations or space-time codes and constellations has been studied widely using both BICM and MLC methods. For example, the coded-modulation design for turbo codes has been reported in [96–98] and for LDPC is described in [22, 99–101]. Polar codes have also been used both with BICM and MLC/MSD. Designing BIPCM has been the subject of much research, e.g., [24, 102, 103] while the best interleaver design has remained an open

problem. In contrast, MLC/MSD with SPM has been employed for constructing polar coded-modulation [25, 26, 67, 104] and due to the clarity of design and the conceptual similarity to channel polarization, MLPCM outperforms BIPCM [25]. Therefore, despite BICM became *a de facto* for wireless systems, e.g., LTE-A [34] due to less decoding delay and simpler design, MLC/MSD has been the dominant candidate in terms of error rate and LLR calculation complexity for designing polar coded-modulation.

To design MLPCM, Trifonov in [92] suggested using the GA, but the full steps of design have not been presented in the literature. Seidl et al. in [25] used the GA to design MLPCM by simulation-based estimating the capacity of each binary channel of a constellation to apply the GA. In this chapter, we explain full steps of the MLPCM and BIPCM design using the GA based on a simple method of estimating the average LLR for QAM instead of using the capacity of each binary channel in an AWGN channel. We also show that the GA can be used to design codes for the orthogonal transmission (e.g., OSTBCs) in slow fading channels when the CSIR and the channel statistics are available. In the GA-based design for slow fading channels, we find the average bit-channel error rates by constructing the code for a small number of fading realizations using GA and choose bit-channels with the lowest average error as the information set.

In practice, the LLR distribution of the subsets of each level of MLC/MSD in most STBCs (other than OSTBCs) are typically far from Gaussian and thus, the GA cannot be employed for the code design. Fortunately, simulation-based design can be implemented with reasonable complexity for very large output alphabet sizes when the FER is not too low. Furthermore, the simulation-based design can approximate the FER during the design procedure. Throughout this thesis, we construct polar codes for most STBCs using the simulation-based design method.

In this chapter, we aim to design the FER minimizing polar codes. These codes are

typically used when the CSIT is not available but the CSIR and channel statistics are known. To this end, we first review the simulation-based design of polar codes for the BPSK and irregular space-time constellations in any arbitrary channel with MLPCM and BIPCM. We then explain the simplified steps and a set of algorithms to design polar coded-modulation using GA for QAM in an AWGN channel. In addition, we propose a method to design the polar coded-modulation that substantially reduces the space and arithmetic complexity compared to GA. Furthermore, we propose a GA-based method and a low-memory-space/low-arithmetic-complexity method to design polar coded-modulation for QAM in slow fading channels. Code design methods for the AWGN channel can also be employed for small to large MIMO systems if the channel is convertible to a set of AWGN channels. However, methods proposed for the slow fading channels can only be used for small MIMO schemes. Note that in this chapter, only low-memory-space/low-arithmetic-complexity methods can be used for the real-time generation of polar codes.

The rest of the chapter is organized as follows: Polar code design based on the density evolution is reviewed in Section 4.1. The simulation-based design for polar coded-modulation is reviewed in Section 4.2; the GA-based design algorithms for an AWGN channel is explained in Section 4.3; a novel low-memory-space/low-complexity polar coded-modulation design for AWGN channels is explained in Section 4.4; the coded-modulation design procedure is described in Section 4.5; a performance comparison of BICM and MLC/MSD for polar coded-modulation is provided in Section 4.6; a low-complexity GA-based method to design polar coded-modulation in slow fading channels is proposed in Section 4.7; a low-memory-space/low-complexity method for designing polar coded-modulation in slow fading channel is proposed in Section 4.8; finally, the conclusions are presented in Section 4.9.

## 4.1 Design Using Density Evolution for BPSK

As an important property of polar codes, the polar graph is fixed. Therefore, density evolution can be employed to design the code. In practice, the numerical implementation of the density evolution is expensive. However, the general description of density evolution helps to understand the aspects of code design. In this section, we describe the density evolution.

In density evolution, the received LLR distribution is evolved through the polar code graph in Fig. 2.4 and the LLR distribution for all nodes in polar code graph is determined. This can be done using density evolution updating rules in [90, Equ. 2] and numerical methods. LLR densities for a polar code with $N = 8$ in an AWGN channel with BPSK are shown in Fig 4.1. In this figure, $\nu_{1:8,1}$ corresponds to nodes 1 to 8 at the left most of the decoder graph in Fig. 2.4. Observe that for nodes $\nu_{1:8,1}$, the LLR is accurately Gaussian. In the second level of the graph, updating using density evolution operators results in Gaussian LLR for nodes $\nu_{1:4,2}$ but it degrades from Gaussian for nodes $\nu_{5:8,2}$. The same procedure is repeated until we find the LLR densities for all bit-channels.

Using the LLR distributions, the mutual information or the tail error probability of bit-channels (i.e., $\mathrm{P}(\lambda_{i,n+1} < 0)$, assuming all zeros codeword is transmitted) can be estimated. Finally, we choose $K$ bit-channels with the highest mutual information or the lowest tail error probability as the information set. In Fig 4.1, the empirical LLR densities of nodes is compared to a Gaussian density with the same mean and variance. Observe that the Gaussian distribution can approximate the empirical distribution. As we will discuss in Section 4.3, this property is used to suggest a simple approach for the code design in AWGN channel. Finally note that since the LLR distribution of nodes depends on SNR and channel statistics, the information set changes with channel statistics and SNR.

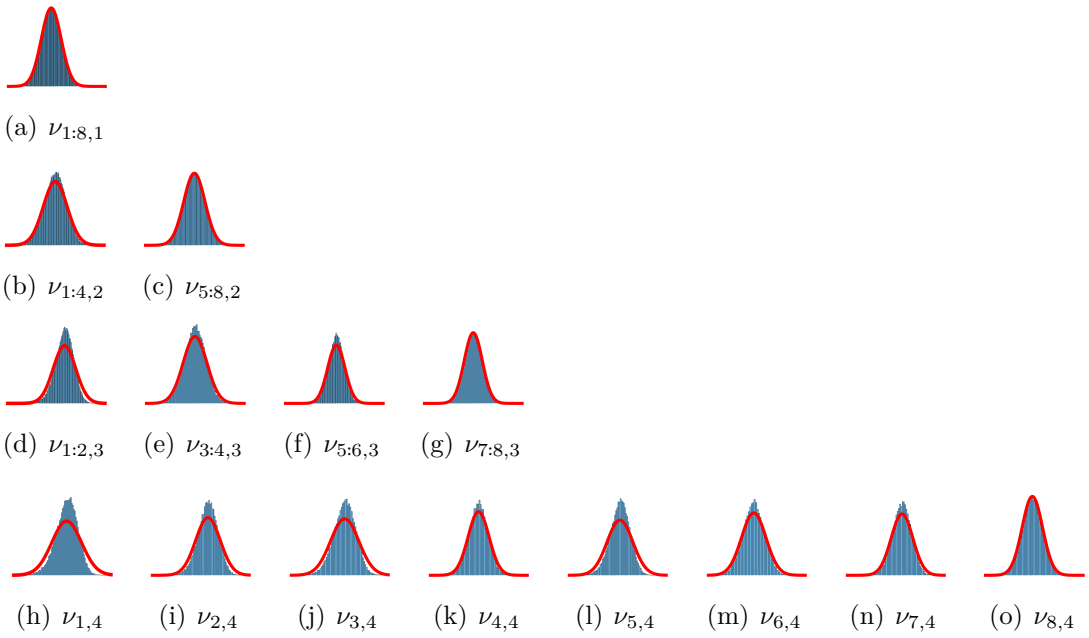



**Figure 4.1:** Comparison of the empirical bit-channels LLR distributions (in blue) and Gaussian distribution (in red) with BPSK in an AWGN channel at $\gamma = 10$ dB.

## 4.2 Simulation-based Design

Among different design methods, simulation-based polar code design benefits from high flexibility for adapting to a variety of practical channels [105]. In this section, we review the simulation-based polar code design for BPSK, MLC/MSD, and BICM.

### 4.2.1 Design for BPSK

To design polar codes, the positions of the information bits must be determined. Determining the information set by using Monte Carlo simulation, proposed by Arıkan in [1], is one of the methods of polar code design which benefits from high flexibility for adapting to a variety of practical channels. In the simulation-based design method, as described in [52], the transmission of a large number of message words is simulated and SCD decodes bits subsequently from the first to the last. Then, the number of

the first error events[7] for each bit-channel is measured. The number of transmitted codewords for achieving the sufficient statistic can be decreased if, after recording each first error event, the corresponding bit is corrected to prevent propagating that error and the next bit-channels are examined subsequently. When the polar code is designed for a predetermined rate $R$ at a specific SNR, the best information set is chosen to minimize the FER by finding the $K$ message bit positions with the lowest number of the first error events.

By recording the position of each first error event for each simulated codeword, it is easy to evaluate the FER for any information set. Any simulated codeword with at least one first error event in positions specified by that information set would also have been decoded incorrectly by a real decoder for the polar code defined by that information set. Thus, the FER for a given information set can be approximated by dividing the number of incorrectly decoded codewords by the number of simulated codewords.

More formally, suppose we simulate $N_{\text{SIM}}$ codewords of length $N$ at a given SNR. Let $\sigma_{n_M,\kappa} = 1$ if a first error event occurred in the $\kappa^{th}$ bit-channel during the $n_M^{th}$ simulated codeword transmission, and $\sigma_{n_M,\kappa} = 0$ otherwise[8]. For a given hypothetical information set, $\mathcal{A}$, the $n_M^{th}$ simulated codeword would have been decoded incorrectly if $\Sigma_{\kappa\epsilon\mathcal{A}}\,\sigma_{n_M,\kappa} > 0$. Let $\iota_{n_M} = 1$ if $\Sigma_{\kappa\epsilon\mathcal{A}}\,\sigma_{n_M,\kappa} > 0$. Out of the $N_{\text{SIM}}$ simulated codewords, the number of codewords that would have been incorrectly decoded is $\Sigma_{n_M=1}^{N_{\text{SIM}}}\iota_{n_M}$.

Obviously, this approach is a Monte Carlo based method and therefore, if a sufficiently large number of frames is used, the designed codes tend to the optimum solution of this method. As a simple algorithm to determine the number of frames that should be used for the simulation-based code design, the FER variation of the

[7]For each codeword, the first error event defined as the first erroneous output bit. This error does not include the propagated error and just represents the error of the bit-channel.

[8]Let $N_{ERR,\kappa} = \Sigma_{n_M=1}^{N_{\text{SIM}}}\sigma_{n_M,\kappa}$ be the total number of first error events in the $\kappa^{th}$ bit-channel. The information set, $\mathcal{A}$, of the minimum-FER code of rate $K/N$ contains the values of $\kappa$ with the $K$ smallest values of $N_{ERR,\kappa}$.

simulation-based code design method is measured by changing the number of frames step-wisely. Once the FER variation remains less than a sufficiently small threshold, the method can be stopped.

### 4.2.2 Design for MLC/MSD

For designing a MLC scheme, $B$ binary codes with the set of rates $\{R_1, R_2, ..., R_B\}$ with each component code length of $N$ should be constructed. The well-known methods of determining the component code rates are capacity rule, balanced distances rule, coding exponent rule, the cutoff rate rule and the equal error probability rule. In the aforementioned methods, the component code rates are determined according to the subchannel capacities, the intra-subset distances of the constellation and the minimum Hamming distance of component codes, the random coding exponent of sub-channels, the cutoff rates of sub-channels and the equality of error probabilities or bounds on error probabilities of subchannels, respectively. The results of all these methods are usually the same. A detailed analysis of these methods has been provided in [106]. However, as the main shortcoming of these methods, the rates are determined based on bounds on the error probability or capacity rather than the actual error rate of the decoder. The capacity rule, as a widely used design rule, can predict approximately correct rates at the SNR corresponding to $\mathcal{I}(\mathbf{Y};\mathbf{S}) = R_{tot}B$ while in most other SNRs it does not provide a good set of rates. Other rules such as the equal error probability rule or the cutoff rate rule have the same problem. However, in the simulation-based code design method, since the practical decoder is employed, rates close to optimum for that particular decoder can be determined. Here, we employ a simulation-based method to design the multilevel polar coded-modulation scheme.

The simulation-based method explained in Section 4.2.1 can be employed to design polar codes for MLC/MSD. In this method, proposed in [52], the number of the first error events of all bit-channels in all levels is evaluated and the $K$ bit-channels with the

lowest number of the first error event among all levels are chosen for the information set. As an example, when we design codes using the capacity rule for 16-QAM and $R_{tot}$ = 0.5, the rates at $E_b/N_0$ = 4 dB in AWGN are {0.099,0.514,0.833,0.991}, while if we use the simulation-based design, the rates are {0.035,0.327,0.681,0.957}. Once we use the capacity rule since the SCD is not an ML decoder, the code rates are more than the actual capacity of subchannels and therefore, it shows a degraded performance in comparison to simulation-based design. This method can successfully design polar codes that work better than turbo code and DVB-S2 LDPC code in a noncoherent system [52].

### 4.2.3 Design for BICM

The simulation-based design method mentioned in Section 4.2.1 can be employed to design polar coded-modulation with BICM [25]. Here, to determine the information set, the number of the first error events is measured based on the direct evaluation of the bit-channels in the simulation of the coded-modulation scheme and the bit-channels with the lowest number of the first error event are selected.

## 4.3 GA-based Design for the AWGN Channel

As shown in [92], polar codes can be designed with low-complexity using the GA of the density evolution in an AWGN channel for a BPSK. In this section, we review the FER-minimizing polar code design using GA for BPSK and we explain the complete steps of MLPCM and BIPCM design in an AWGN channel using the simplified LLR estimation for QAM introduced in Section 3.1.1.

### 4.3.1 Design for BPSK

In GA, the LLR distributions for each node in the polar graph can be approximated with a set of Gaussian distributions, in which the variance of the distribution $\sigma^2$ is two times the average LLR of a specific bit-channel. Therefore, to implement the GA-based design for polar codes we only need to update the average LLR through the polar graph. The updating average LLR rule for upper bit-channels is $\phi^{-1}(1-[1-\phi(\bar{\lambda}_1)][1-\phi(\bar{\lambda}_2)])$ and for lower bit-channels is $\bar{\lambda}_1+\bar{\lambda}_2$ where $\bar{\lambda}=E[\lambda]$ [93]. From [107], $\phi(x)$ is approximated as

$$\phi(x) \triangleq \begin{cases} 1 & x=0, \\ \exp(-0.4527x^{0.86}+0.0218) & 0<x\leq 10, \\ \sqrt{\frac{\pi}{2}}(1-\frac{10}{7x})\exp(\frac{-x}{4}) & 10<x. \end{cases} \tag{4.1}$$

The GA-based method for finding the information set is described in Algorithm 2. In this algorithm, we estimate the conditional error probability of the $i^{th}$ bit-channel as $v_i = Q(\sqrt{\bar{\lambda}_{i,n+1}/2})$ [92, 108], given that the previous bit-channels are known, using Function GA-BER and sort out $K$ bit-channels with the lowest error rates to determine the information set.

**Algorithm 2:** GA-based Design for BPSK

**Input :** $\gamma$= SNR [dB]
**Output:** The sorted bit-channel indices **idx**
**Procedures:**
[**v**,**idx**] =Sort(**v**): Sorts the input vector in an increasing order and outputs the sorted indices **idx** and sorted values **v**.

1 $\bar{\lambda}=4\times 10^{(\frac{\gamma}{10})}$
2 **v**=GA-BER($\bar{\lambda}$,N)
3 [**idx**]=Sort(**v**)
4 return **idx**

**Function** GA-BER($\bar{\lambda}$,$N$) (From [93] with modification)

**Input:** $\bar{\lambda}$ and $N$
**Output:** The vector of the BER of bit-channels given that the previous bit-channels are frozen $\mathbf{v}$

$\bar{\lambda}_{1,1} = \bar{\lambda}$
**for** $j = 1 : \log_2 N$ **do** ▷ Estimating $\lambda$ for $(j+1)^{th}$ stage.
  **for** $i = 1 : 2^{j-1}$ **do**
    $\bar{\lambda}_{2i-1,j+1} = \phi^{-1}(1 - [1 - \phi(\bar{\lambda}_{i,j})]^2)$ ▷ Estimating $\lambda$ for the $i^{th}$ bit-channel.
    $\bar{\lambda}_{2i,j+1} = 2\bar{\lambda}_{i,j}$
**for** $\kappa = 1 : N$ **do**
  $v_\kappa = Q\left(\sqrt{\frac{\bar{\lambda}_{\kappa,\log_2(N)+1}}{2}}\right)$
return $\mathbf{v}$

The FER of the polar coded scheme for BPSK in an AWGN channel can be approximated as [25]

$$P_K = 1 - \prod_{i=1}^{K}(1 - v_i), \tag{4.2}$$

### 4.3.2 Design for MLC/MSD

To construct polar codes using GA, the average LLR for each binary channel of a QAM constellation is estimated and an independent binary polar code is constructed for each level. Due to the Gaussian noise, the LLR distribution for a BPSK constellation in an AWGN channel is $\mathcal{N}(\frac{4}{N_0}, \frac{8}{N_0})$. However, the LLR distributions of each level of a multilevel coding scheme for PAM and QAM constellations are not known. The average LLR of QAM with $2^B$ points can be estimated using the average LLRs of the constituent PAM with $2^{B/2}$ points based on the transform described in Theorem 1. In this theorem, we indeed use a linear block code to estimate the LLRs of QAM with 2D SPM from LLRs of PAM. This linear block code is similar to the second stage of a polar encoder and hence, a modified GA can be employed to estimate the average LLRs of QAM. The average LLR estimation algorithm is formally described

in Function AverageLLR. To estimate the average LLRs of a PAM constellation, we use the piecewise linear approximation in Theorem 2. Consequently, since the LLR is a piecewise linear function of $y$ and the noise is Gaussian, the average LLR of each level of PAM, given a zero is transmitted, can be computed through integration. For the first level of a PAM, the integral is given by

$$\bar{\lambda}_1 = \sum_{d|s_d \in \mathcal{X}_1^0} \int_{\Omega_d} \mathrm{P}(s_d)\lambda_1(y) \frac{1}{\sqrt{\pi N_0}} e^{\frac{-(y-s_d)^2}{N_0}} dy, \tag{4.3}$$

where $\mathrm{P}(s_d)$ is the probability of the transmission of $s_d$ within $\mathcal{X}_1^0$ and $\Omega_d$ is defined in Theorem 2. For example for 4-PAM, $\mathrm{P}(s_d) = 0.5$ and for 8-PAM, $\mathrm{P}(s_d) = 0.25$. For other levels, the integration is taken over the corresponding set $\mathcal{X}_b^0$. Therefore, the general form of (4.3) can be written as

$$\bar{\lambda}_b = \sum_{d|s_d \in \mathcal{X}_b^0} \int_{\Omega_d} \mathrm{P}(s_d)\lambda_b(y) \frac{1}{\sqrt{\pi N_0}} e^{\frac{-(y-s_d)^2}{N_0}} dy, \tag{4.4}$$

where $\lambda_b(y)$ is given by the LLR approximation in (3.11). As a numerical example, for 4-PAM and 8-PAM, the average LLRs of the binary channels are $[6.3, 31.9]$ and $[0.7, 6.0, 30.5]$ at $\gamma = 10$ dB, respectively.

**Function** AverageLLR($\bar{\boldsymbol{\lambda}}$)

**Input:** $\bar{\boldsymbol{\lambda}}_{\mathrm{PAM}}$: The $\bar{\boldsymbol{\lambda}}$ of PAM with SPM in (4.4)
**Output:** $\bar{\boldsymbol{\lambda}}_{QAM}$: The vector of average LLRs of $2^B$-QAM with SPM

**for** $\kappa = 1 : \frac{B}{2}$ **do**
  $\bar{\lambda}_{QAM,2\kappa-1} = \phi^{-1}(1 - [1 - \phi(\bar{\lambda}_{PAM,\kappa})]^2)$
  $\bar{\lambda}_{QAM,2\kappa} = 2\bar{\lambda}_{PAM,\kappa}$
return $\bar{\boldsymbol{\lambda}}_{QAM}$

In [92], it is noted that the FER of MLPCM can be approximated with $1-\prod_{b=1}^{B}(1-P_b)$ where $P_b$ is the FER of the $b^{th}$ level. By extending this bound, the total FER of

MLPCM can be approximated as [25]

$$P_K = 1 - \prod_{b=1}^{B} \prod_{i=1}^{K_b} (1 - v_b^i), \tag{4.5}$$

where $K_b$ is the message length of the $b^{th}$ level and $v_b^i$ is the BER of the $i^{th}$ bit-channel in $b^{th}$ level given previous bit-channels are frozen. To find component code rates, [92] suggests using the equal error probability rule in which all levels of a MLPCM have approximately the same FER. However, this requires solving a program to find code rates. From (4.5), one can observe that MLPCM works like a longer single binary polar code while it observes a variety of equivalent SNRs corresponding to the different levels. Thus, to determine component code rates, the bit-channel reliabilities $v_b^i, \forall i = 1, ..., N$ can be measured for all levels $b = 1, ..., B$ and among them, those with the lowest genie-aided BERs are chosen as the total information set [25]. This automatically determines the rate of each level since some bit-channels of each level are in the total information set. When we use this rule to design the code, the FERs of different levels are very close. However, this rule does not require solving any program to determine the code rates which highly simplifies the MLPCM design. The entire code design procedure for MLPCM is mentioned in Algorithm 3. The procedure Sort used in the algorithm is defined in Algorithm 2. The output of the algorithm is the ordered bit channels, $\mathbf{idx}_{tot}$ that we choose the first $K$ as the information set. Note that we finally only return the total information set as it is enough for the encoding and decoding processes and component code rates should not be used necessarily.

Determining component code rates for the sake of the simpler implementation or the research might be of interest. Here, we present a simple method for determining

**Algorithm 3:** GA-based design for MLC/MSD with QAM

**Input :** The average LLR vector for PAM $\bar{\boldsymbol{\lambda}}_{PAM}$, $B$, and $N$
**Output:** The sorted bit-channel indices $\mathbf{idx}_{tot}$

$\bar{\boldsymbol{\lambda}}_{QAM}$=AverageLLR($\bar{\boldsymbol{\lambda}}_{PAM}$)
**for** $b = 1 : B$ **do**
  $\mathbf{v}_b$=GA-BER($\bar{\lambda}_{QAM,b}$,$N$) ▷ Determining conditional BERs of each level
[.,$\mathbf{idx}_{tot}$]=Sort($[\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_B]$) ▷ Finding the total information set
return $\mathbf{idx}_{tot}$

component code rates described in Algorithm 4. In this method, we find the labels of bit-channels of each component code by searching the ordered bit channels, $\mathbf{idx}_{tot}$, found using Algorithm 3.

**Algorithm 4:** Determining component code rates

**Input :** $R$, $B$, $N$, and $\mathbf{idx}_{tot}$
**Output:** $\mathbf{R}$
**Procedures:**
Find($\mathbf{v}, x_1, x_2$): Finds the number of positions in $\mathbf{v}$ with values in the range of $[x_1, x_2]$.

$K = \lfloor RBN \rfloor$ ▷ Finding the total message length
$\mathcal{A} = \mathbf{idx}_{tot,1:K}$
**for** $b = 1 : B$ **do**
  $k_b = \text{Find}(\mathcal{A}, (b-1)N, bN)$
  $R_b = \dfrac{k_b}{N}$
return $\mathbf{R}$

### 4.3.3 Design for BICM

In GA the distribution of the decoder input LLRs, $\lambda_{i,1}$, should be close to Gaussian. The distribution of input LLRs for MLPCM with a SPM and BIPCM with Gray mapping in an AWGN channel for a 16-QAM constellation is shown in Fig. 4.2. It is clear that while the LLR distribution of binary channels in MLC/MSD is close

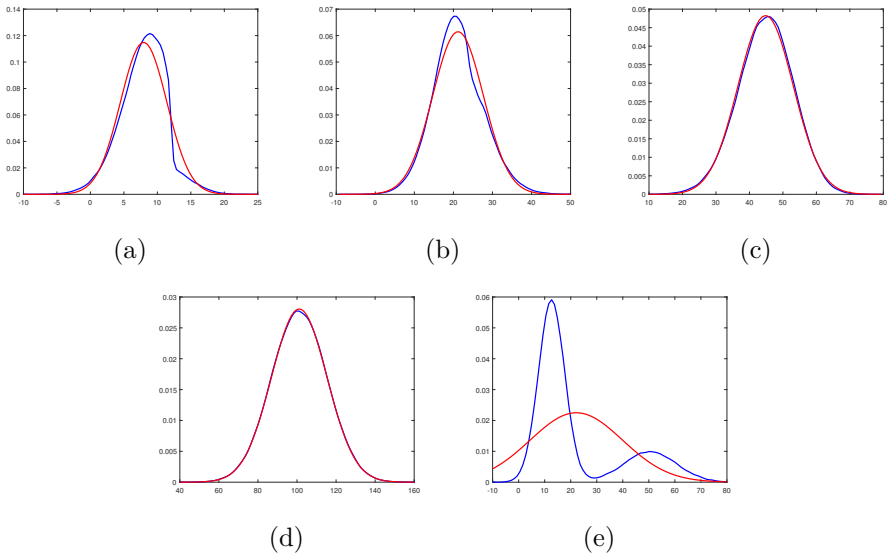



**Figure 4.2:** Comparison of the empirical LLR distribution (blue) and an approximated Gaussian distribution (red) of 16-QAM in an AWGN channel at $\gamma = 15$ dB for MLC/MSD binary channels a) 1, b) 2, c) 3, d) 4, and e) for BICM.

to Gaussian, the LLR distribution of BICM is far from the Gaussian. Thus, GA is not suitable for designing BIPCM and throughout this thesis, we design our codes using the simulation-based method. This also has been observed in [104]. In fact, we observe that if the input LLR distributions are relatively close to Gaussian, e.g., they have only one peak, the GA method can generate codes that are as good as the simulation-based method.

At low SNRs, the LLR distribution of BICM is not very far from Gaussian and acceptable polar codes can be constructed using GA. In those cases, the procedure is the same as MLC/MSD in which we first estimate the average LLRs and then we design the polar codes using the GA method similar to BPSK case in Section 4.3.1. As shown in [104], the FER of codes designed using this method is slightly worse than the simulation-based method.

## 4.4 An Efficient Design Method for the AWGN Channel

Due to the simple rate matching, the polar code rate can be changed quickly. Thus, low-complexity low-memory-space constructing of polar codes can result in more applicability. In this section, we propose a novel design approach for low-complexity low-memory-space construction of FER-minimizing polar codes. This method can also be employed for designing throughput optimal polar coded-modulation introduced in Chapter 5 for AMC.

The bit-channel error rates, $v_b^i$, can be approximated using the GA method or the simulation-based method at a wide range of SNRs. To achieve the best bit-channels, we can store these values or the information set of the corresponding constructed polar code at different SNRs. This needs a large space. Instead, we can model $v_b^i$ at all SNRs with only a few parameters and choose codes by regenerating the bit-channel error and selecting the best bit-channels at any target SNR. This requires much lower space. For example when $N = 1024$ and $R_{tot} = 1/2$, for storing the information set we need 5120 bits at each SNR. If we store them for 60 SNRs, they would occupy around 38 kilobytes (KB). We also can store bit-channels in a mask by setting zero for frozen and one for information bits. In this case, the total memory of 8 KB is needed. Instead, if we model each bit channel with 3 parameters, we only need around 3 KB assuming 1 byte is enough for storing each number. The complexity of one time running of GA is $O(BN)$ while the complexity of the proposed method is $O(BN)$ with much lower number of multiplications. Here, we propose to model the $i^{th}$ bit-channel error rate as

$$v^i(\gamma) \approx 0.5Q(\vartheta_1^i 10^{\frac{\vartheta_2^i \gamma}{20}} + \vartheta_3^i). \tag{4.6}$$

In (4.6), $\vartheta_1^i$, $\vartheta_2^i$, and $\vartheta_3^i$ are determined using the least squares fitting. We chose this metric since the Q-function models the BER for a BPSK in an AWGN channel.

To design MLPCM using this method, the average LLR of each level is estimated and converted to the SNR. Then, based on (4.6), the bit-channels errors are generated and bit-channels with the lowest error are selected as the information. The codes generated using this method perform approximately the same as the codes generated using the GA method.

Due to the simple rate matching, polar code rate can be changed fast based on the requirements of the system or the use of AMC. Thus, due to the low memory-space and arithmetic-complexity requirements of the proposed method, it can be implemented for the real-time polar code-modulation design. Note that the low-complexity polar code design by considering the partial order of bit-channels, introduced in [109], can be used with the proposed method for further reducing the space complexity.

## 4.5 Coded-Modulation Design Procedure

For designing polar coded-modulation schemes with either MLC or BICM, first we optimize the constellation, we then find a good bit-to-symbol mapping using the BICM or MLC labeling algorithm for the optimized constellation, and we finally design the polar code given the constellation and the bit-to-symbol mapping. A random interleaver is used to construct BICM with a length of $N_{tot}$. For designing the polar code given a fixed FER, the bisection design SNR search algorithm in [52, Algorithm 2] can be employed.

## 4.6 Comparison of Polar Coded-Modulation with BICM and MLC/MSD

In Section 3.2, we showed the gap between the BICM capacity and that of the constellation-constraint capacity for 16-QAM. This gap exists for almost all constellations other than BPSK and quadrature phase shift keying (QPSK). We also explained in Section 3.1 that the decoding procedure in MLC/MSD suffers from a performance loss since it is not an ML decoder. As explained in Section 3.2, BICM decoder is also suboptimal. Therefore, there might be trade-offs in design and utilization of these schemes that may result in the superiority of one of them in comparison to another in different situations. In this section, we review some basic results for an AWGN channel with 2D constellations and in next chapters, we compare the polar coded-modulation construction based on BICM and MLC/MSD for a variety of systems and channels. The total code rate and the total code length for all curves of this section are 0.5 and 4096 bits, respectively. For constructing BICM and MLC/MSD schemes, Gray mapping and SPM are employed, respectively. For all cases, the codes are designed at a FER of 0.01 using the simulation-based method. The SCLD decoder uses CRC with $L_{CRC}$ = 16 bits.

Fig. 4.3 shows the FER of polar coded-modulation constructed using BICM and MLC/MSD with QPSK. BICM and MLC/MSD show almost the same performance for both SCD and SCLD since in both cases the generated LLRs are the same. The same comparison has been provided for a 16-QAM constellation in Fig. 4.4. Here, we observe that MLC/MSD as reported in [67, 104] outperforms the BICM scheme for both SCD and SCLD. As a trade-off, MLC/MSD due to the successive decoding of levels has a slightly longer decoding delay in comparison to BICM that may not be suitable for some delay-sensitive applications. In contrast, the LLR calculation in the decoding of BICM has more complexity than the MLC/MSD scheme since in

MLC/MSD scheme using the information of upper levels, the complexity of computing LLRs in lower levels can be substantially reduced.

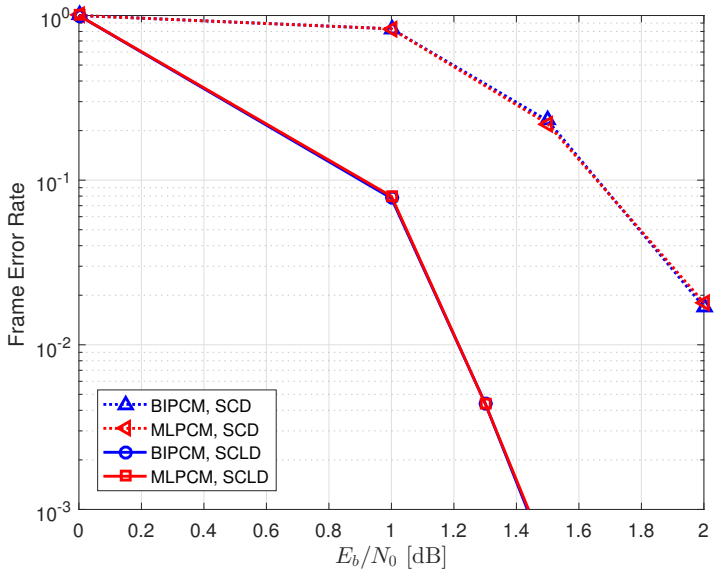



**Figure 4.3:** Comparison of FER of polar coded-modulation schemes constructed using BICM and MLC/MSD for a QPSK constellation in an AWGN channel.

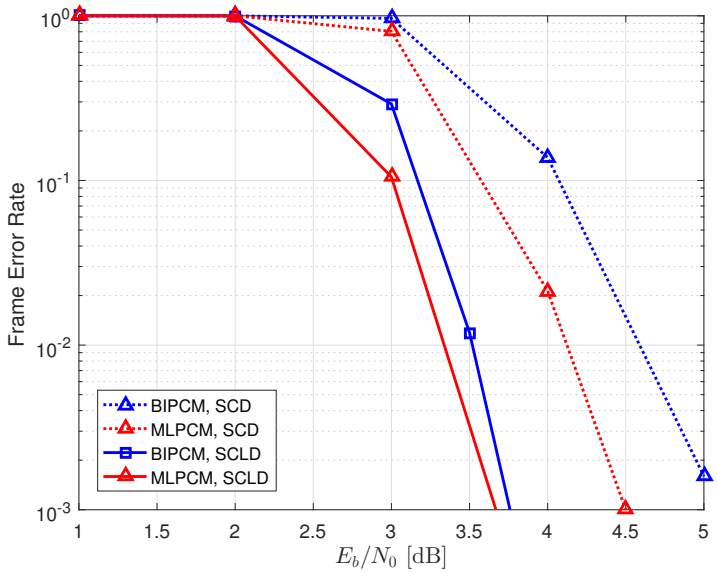



**Figure 4.4:** Comparison of FER of polar coded-modulation schemes constructed using BICM and MLC/MSD for a 16-QAM constellation in an AWGN channel.

## 4.7 A Novel Design Method for Slow Fading Channels

When the CSIT is not available but the fading statistics are available, the FEC code should be designed to work well on average for all fading realizations. Once more, the simulation-based polar code design can be used to find the information set but due to its complexity when we use large codes or when we design the code at low FERs, proposing alternative methods with lower complexity is of importance. In this section, we propose novel methods for designing MLPCM in a slow fading channel based on the GA.

When QAM is used in an AWGN channel, the LLR distributions of MLC/MSD at a wide range of SNRs can be approximated with a set of Gaussian distributions. Thus, we can use the GA to design the polar codes for QAM in slow fading channels when an orthogonal transmission method is used, e.g., for single-input multiple-output and OSTBC. However, we cannot employ the GA method for nonorthogonal schemes, e.g., Golden code [35], since LLR distributions are in general far from Gaussian at a wide range of SNRs.

In the first code design method, a set of random realizations of $\mathbf{H}$ is generated and the average LLR of binary channels of the constellation is computed. Then, the conditional BER of bit-channels, $v_b^i$, for each realization of the fading is determined using the GA method. After running this for several fading matrix realizations, we take the average over $v_b^i$, and the information set is determined as the $K$ bit-channels with lowest conditional BERs. This method is described in Algorithm 5. Note that in practice only a few fading matrix realizations is needed to achieve the best code designed using the simulation-based approach. Fig. 4.5 shows the number of fading realizations needed to achieve the lowest average FER code for the Alamouti code [110] with a 16-QAM constellation and $N_{tot}$ = 256 bits. Observe that with

only 35 realizations, we can design the best code found using the simulation-based approach. The complexity of this method can be estimated as $O(BN_FN)$ where $N_F$ is the number of fading realizations used for the code design. This complexity, however, is much less than the complexity of the simulation-based approach since a large number of Monte Carlo iterations (e.g., 100000) are needed to design the code using simulation.

In Algorithm 5, the average LLR of a PAM constellation for each realization of the fading can be estimated similar to the AWGN in 4.4. In this case, the average LLR can be given as

$$\bar{\lambda}_b(\mathbf{H}) = \sum_{d|s_d \epsilon \mathcal{X}_b^0} \int_{\Omega_d} \mathrm{P}(s_d)\lambda_b(y)\mathrm{P}(\mathbf{y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma})dy. \tag{4.7}$$

Note that in 4.7, $\mathrm{P}(\mathbf{y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma})$ can be estimated with a very low-complexity using a simplified decoder explained in Section 6.1. For the sake of approximating the FER of the coded-modulation scheme, (4.5) for slow fading can be modified as

$$P_K = 1 - \frac{1}{N_F} \sum_{n_f} \prod_{b=1}^{B} \prod_{i=1}^{K_b} \left(1 - v_b^i(\mathbf{H}_{n_f})\right). \tag{4.8}$$

To use (4.8) for approximating the FER, first the code is constructed using the Algorithm 5, and for $N_F$ fading realizations, the FER of the information set is estimated and is averaged out. Fig. 4.6 shows the mean square error of the FER approximation for different numbers of fading realizations. We observe that with around 1000 realizations, we have mean square approximation error of $10^{-5}$.

**Algorithm 5:** GA-based design for QAM in slow fading channel

**Input :** $N$, $N_F$, and $B$
**Output:** The sorted bit-channel indices $\mathbf{idx}_{tot}$
**Procedures:**
PAM_OSTBC_AvgLLR($\mathbf{H}$): Estimates the Average LLR of PAM for OSTBCs based on (4.11) for a given fading realization $\mathbf{H}$.
randn($N_t, N_r$): Generates a matrix with random normally distributed elements of size $N_t \times N_r$.

**for** $n_f = 1 : N_F$ **do**
  $\mathbf{H}$=randn($N_t, N_r$)
  $\bar{\boldsymbol{\lambda}}_{\text{PAM}}$=PAM_OSTBC_AvgLLR($\mathbf{H}$)
  $\bar{\boldsymbol{\lambda}}_{\text{QAM}}$=AverageLLR($\bar{\boldsymbol{\lambda}}_{\text{PAM}}$) ▷ Defined in Section 4.3.2.
  **for** $b = 1 : B$ **do**
    $\mathbf{v}_{b,n_f}$=GA-BER($\bar{\lambda}_{\text{QAM},b}$,$N$) ▷ Defined in Section 4.3.1.
**for** $b = 1 : B$ **do**
  $\mathbf{v}_b = \frac{1}{N_F} \sum_{n_f} \mathbf{v}_{b,n_f}$
[$\mathbf{idx}_{tot}$]=Sort([$\mathbf{v}_1$,$\mathbf{v}_2$,...,$\mathbf{v}_B$]) ▷ Defined in Algorithm 2.
return $\mathbf{idx}_{tot}$

## 4.8 An Efficient Design Method for the Slow Fading Channel

The method proposed in Section 4.4 can be employed to design polar coded-modulation with low memory-space and complexity in a slow fading channel. In fact, the GA-based code design for slow fading channels, described in Section 4.7, can be further simplified since instead of evaluation of GA for each fading coefficient, the code can be designed by evaluating the average LLR of each level. In this case, once more the designed MLPCM has the same performance as MLPCM designed using Algorithm 5. This method herein is referred to as the simplified method 1.

For the sake of further simplification, we can directly model the bit-channel average error rates for the slow fading channel instead of the AWGN channel. However, we need to use a different expression to model the error rate of bit-channels. In

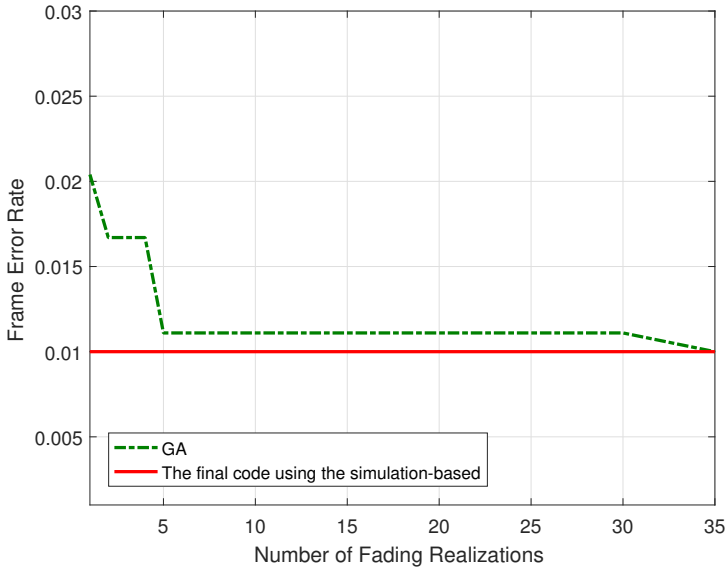



**Figure 4.5:** Evaluation of the effect of number of fading realizations on the FER for code design based on the GA in the slow fading channel for the Alamouti code with 16-QAM and $N_{tot} = 256$ bits.

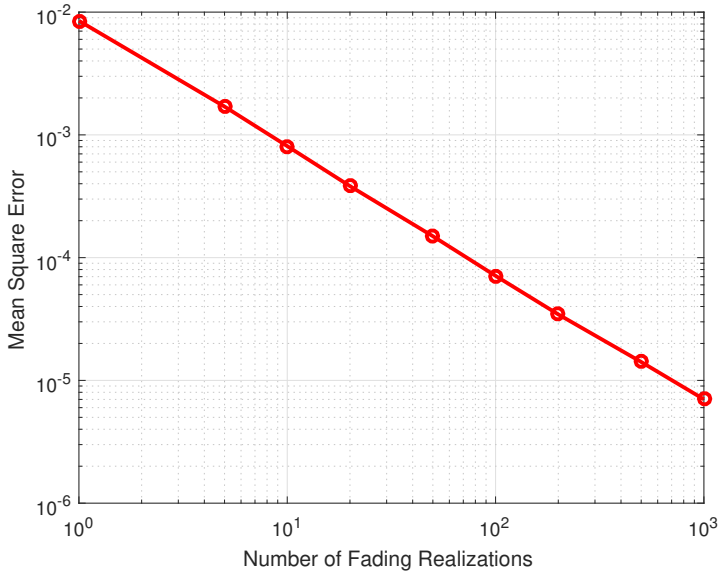



**Figure 4.6:** Mean square error of the FER approximation found using the bound in (4.8) vs. the number of fading realizations for the Alamouti code with 16-QAM and $N_{tot} = 256$ bits.

this case, we use the BER of the maximum ratio combining since it can model the bit-channel error rates. It can be given as

$$v_i(\bar{\gamma}) \approx p_{MRC}(\bar{\gamma}) \sum_{j=0}^{d_s-1} \binom{d_s+j-1}{j} (1-p_{MRC}(\bar{\gamma}))^j, \tag{4.9}$$

where $d_s$ is the order of diversity of the signal and $p_{MRC}$ is given as

$$p_{MRC}(\bar{\gamma}) = \vartheta_1^i - \vartheta_1^i \left(1 + \frac{1}{10^{\frac{\bar{\gamma}}{10}} \vartheta_2^i}\right)^{\vartheta_3^i}. \tag{4.10}$$

In (4.10), $\vartheta_1$, $\vartheta_2$, and $\vartheta_3$ are found using the least squares fitting to bit-channel error rates generated using the GA method described in Section 4.7. To generate codes for MLPCM, the average LLR of each level can be estimated for slow fading channels. It can be given as

$$\bar{\lambda}_b = \sum_{d|s_d \in \mathcal{X}_b^0} \int_{\mathbf{H}} \int_{\Omega_d} \mathrm{P}(s_d) \lambda_b(y) \mathrm{P}(\mathbf{y} \mid \mathbf{S}, \mathbf{H}, \bar{\gamma}) dy d\mathbf{H}. \tag{4.11}$$

Although the method works well for BPSK, the codes generated for MLPCM have higher FER in comparison to codes generated using Section 4.4. In fact, the average LLR or the average SNR of the slow fading is not an appropriate metric for designing component codes and choosing the component code rates. This method herein is referred to as the simplified method 2. The performance of the simplified method 2 is compared with the GA-based method proposed in Section 4.7, the simplified method 1, and the simulation-based approach in Fig. 4.7 for 16-QAM in a slow fading channel with $N_{tot} = 256$ and $R_{tot} = 1/2$. The coded-modulation is designed at the minimum SNR corresponding to a FER of 0.01. Observe that the simplified method 2 is around 0.5 dB worse than the other methods. This corresponds to a space/complexity-performance trade-off in code design methods. Note that the space required for storing the bit-channel model coefficients for the simplified method 2

is $N_F$ times lower than the simpled method 1. In addition, the complexity of the simplified method 2 is $O(BN)$ while the complexity of the simplified method 1 is $O(BN_FN)$ in terms of the number of procedures.

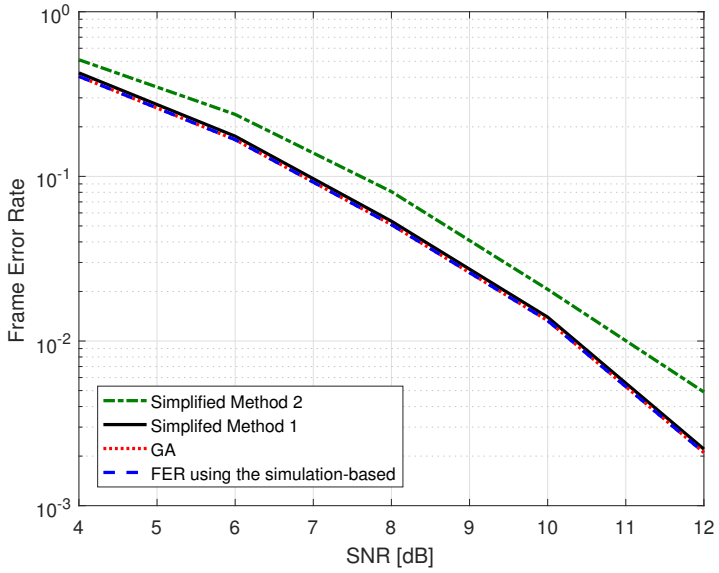



**Figure 4.7:** Comparison of the FER of codes constructed using the proposed simplified method and the best codes found using GA-based method for the Alamouti STBC with 16-QAM and $N_{tot}$ = 256 bits.

## 4.9 Conclusion

Throughout this chapter, we explained a set of simplified steps and algorithms to design polar coded-modulation using GA for QAM in an AWGN channel. We also presented a novel low-complexity method based on the GA to design polar coded-modulation for the orthogonal transmission with QAM in slow fading channels that can generate codes with the same FER as the simulation-based method. Furthermore, we proposed a low-memory-space/low-arithmetic-complexity method to construct polar coded-modulation for QAM in an AWGN channel by modeling the error

rates of bit-channels using an appropriate function. This method has lower complexity compared to GA and can be used for the real-time construction of the polar coded-modulation schemes. The same method can be used in the construction of polar coded-modulation for QAM in slow fading channels. To further decrease the space and the complexity of the code design method, we modeled the error rates of bit-channels for the slow fading channel. However, the FER of the new code design method is worse than GA-based and simulation-based methods. This corresponds to a trade-off between the memory-space and the arithmetic complexity of the code design method and the performance of the code.

# Chapter 5

# Throughput-based Design for Polar Coded-Modulation

The time-varying nature of wireless channels requires the use of AMC schemes to achieve high throughput communication. HARQ error control protocols can enhance the throughput of AMC especially when high order modulation schemes are used. Both AMC and HARQ protocols have been employed widely in communication systems including 4G wireless networks [111]. They are expected to play a central role in 5G and beyond as well, especially in use-cases which require ultra-reliable communication.

Typically, the design objective of binary error correction codes for the AWGN channel is to minimize the error rate, for a given code rate and SNR. This method of design has been widely used for convolutional codes [112], parallel concatenated (turbo) codes [113], LDPC [114] and polar codes [1, 60]. However, to optimize the AMC and HARQ protocols, the throughput is a much more relevant performance metric than error rate, where throughput is defined as the average rate of successful message delivery and indicates how close the performance of a system is to the channel capacity. Designing throughput-optimal codes and coded-modulation typically involves an exhaustive search over a set of code rates and modulations and

employs simulation to estimate the throughput [115]. However, as demonstrated in this chapter, this process can be greatly simplified for polar codes.

When designing polar codes based on the throughput, the optimal set of information bits to maximize the throughput is chosen. The particular advantage of polar codes that facilitates their optimization for maximizing the throughput is this straightforward design method in comparison to most other modern codes. These polar codes are particularly useful for non-combining (NC)[9] and Chase-combining (CC) HARQ where, for the retransmission of a failed codeword, the whole codeword should be retransmitted and incremental redundancy HARQ (IR HARQ) is not employed. Although efficient IR HARQ schemes have been proposed (e.g. [116]) their design and scheduling, especially for coded-modulation schemes, is difficult [117]. However, limited feedback can be employed to achieve high throughput using AMC and HARQ protocols since the polar code graph is fixed and only the information set should be modified when the SNR changes. Polar codes have also been designed to maximize the throughput of CC HARQ with SCD in [118] and with SCLD in [117] based on puncturing for a fixed message length and a BPSK constellation. These methods, given the message length $K$, require a full search over all puncturing lengths from 0 to $N - K$ and thus require running the GA method for $N - K$ times. However, in our proposed code design method, we only need to run the GA method once for most protocols. Therefore, their design is at least $N - K$ times more complex than our method and, as we will show, there is no advantage in using their algorithms. Similar to BPSK modulation, designing MLPCM using the puncturing-based search methods in [117] and [118] is hard since they require a full search over all levels of MLC.

The presented discussions can be extended to feedback assisted MIMO channels. The SVD is a well-known MIMO scheme that decomposes the fading channel into parallel SISO channels with different gains. The MIMO-SVD scheme with optimal

[9]Non-combining HARQ is also known as Type-I HARQ.

power allocation is a special case of optimal precoder and decoder for a MIMO channel [119]. MLPCM can also be designed for MIMO-SVD efficiently.

In this chapter, for the AWGN channel, a set of algorithms for designing MLPCM using GA is proposed that is based on maximizing the throughput, instead of the well-studied objective of minimizing the FER. In these algorithms, we fix the code-length and find the optimal information set using bounds on the throughput of the coded-modulation schemes. These codes are designed for NC and CC HARQ schemes with SCD. In addition, since polar codes optimized for SCD are suboptimal for SCLD, a fast rate matching algorithm is proposed to find the code rate corresponding to the maximum throughput for SCLD when used with polar codes designed for SCD. For MIMO fading channel, polar codes are designed based on the throughput for MIMO-SVD scheme.

In particular, first, we show that by adapting MLPCM to SNR, throughput very close to the capacity can be achieved. In this case, CC HARQ does not provide any advantage over NC HARQ. However, when the transmitter is restricted to use a smaller number of codes, CC HARQ outperforms NC HARQ and also provides higher throughput in comparison to BICM-based CC and IR HARQ, constructed in [120]. We also show that when the levels of MLPCM can be decoded independently using HARQ, the throughput is enhanced substantially.

Methods introduced in this chapter can be used for the AWGN channel. However, due to the expensive computation of SVD for a large number of antennas, the SVD can only be used for small MIMO systems. The throughput-based design of polar codes can be implemented for both offline and real-time code design.

The rest of the chapter is organized as follows: The system model and HARQ protocols are described in Sections 5.1 and 5.2, respectively. The throughput of HARQ protocols as a design metric is explained in Section 5.3. The polar code design methods based on the throughput for a BPSK constellation with SCD are introduced

in Section 5.4, and the MLPCM design procedure for the QAM constellations is described in Section 5.5. In addition, a rate matching algorithm for SCLD is proposed in Section 5.6. Furthermore, a MIMO-SVD scheme is proposed in Section 5.7. Finally, numerical results and discussions are provided in Section 5.8, and conclusions are presented in Section 5.9.

## 5.1 System Model

In this chapter, we follow the system model defined in Chapter 2 by employing AMC and HARQ protocols. For HARQ systems, the transmitter adds a cyclic redundancy check (CRC) sequence of length $L_{\text{CRC}}$ to data to generate a repeat request when the codeword is not received correctly. For HARQ protocols used in the AWGN channel, the system model can be written as

$$y_{i,l} = s_{i,l} + w_{i,l}, \tag{5.1}$$

where $y_{i,l}$ is the $i^{th}$ received sample in the $l^{th}$ retransmission and $s_{i,l}$ and $w_{i,l}$ represent the transmitted symbol and the noise, respectively. The LLR calculation at each level can be written as

$$\lambda_{b,i,l} = \ln \frac{\sum_{s \in \mathcal{X}_b^0} \mathrm{P}(y_{i,l} \mid s_{i,l} = s)}{\sum_{s \in \mathcal{X}_b^1} \mathrm{P}(y_{i,l} \mid s_{i,l} = s)}, \tag{5.2}$$

## 5.2 HARQ Protocols

In this chapter, we design MLPCM for four HARQ protocols. The protocols are divided into two groups, level-dependent and level-independent, based on the dependency of levels for decoding.

**Level-Dependent HARQ Protocols**

In level-dependent protocols, all levels of a multilevel codeword are decoded and

an ACK is fed back to the transmitter only if all levels are correct. Indeed, all levels of the multilevel code are dependent and all levels are seen as one codeword. Thus, only one CRC for checking the correctness of the codeword is employed.

The LLRs used for each decoding attempt depend on whether NC or CC is used. For NC level-dependent (NC-D) HARQ, the LLRs depend only on the received sample for the latest retransmission of the codeword, so the LLRs are given by (5.2). For CC level-dependent (CC-D) HARQ, the LLRs depend on the received samples from all retransmissions, according to

$$\lambda_{b,i,L} = \ln \frac{\sum_{s \in \mathcal{X}_b^0} \prod_{l=1}^{L} \mathrm{P}(y_{i,l} \mid s_{i,l} = s)}{\sum_{s \in \mathcal{X}_b^1} \prod_{l=1}^{L} \mathrm{P}(y_{i,l} \mid s_{i,l} = s)}, \tag{5.3}$$

where $L$ is the number of retransmissions.

**Level-Independent HARQ Protocols**

In level-independent protocols, as proposed in [121], independent codewords with their own CRC are transmitted on each level of MLC. For each new codeword, the receiver decodes the codeword of each level independently and checks whether it is a valid codeword. When a codeword of a specific level is invalid, the same codeword is retransmitted during the next transmission on the same level while on all other upper and lower levels, new codewords containing new messages are transmitted. In this protocol, the receiver waits until a codeword of a level is decoded correctly before attempting to decode the next levels.

An example of the protocol is shown in Fig. 5.1. In this example, codewords $A_1$, and $B_1$, are transmitted using the first and second levels of a two-level MLC, and codeword $A_1$ is decoded incorrectly. No attempt is made to decode $B_1$, (since its LLRs cannot be calculated without reliable knowledge of $A_1$), so a NACK is sent to the transmitter informing that $A_1$ failed. The transmitter responds by retransmitting $A_1$ on level 1, and transmitting a new level-2 codeword, $B_2$, on level 2. In this example

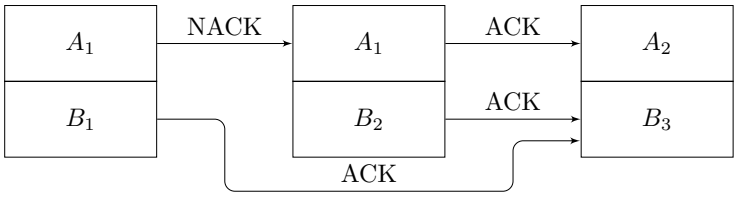



**Figure 5.1:** Illustrative example of the use of level-independent HARQ.

the receiver successfully decodes $A_1$ after this transmission, so now it can attempt to decode $B_1$ (using the received samples from the previous transmission) and $B_2$ (using the current received samples). If $B_1$ fails, the transmitter is instructed to send $A_2$ and $B_1$. If $B_2$ fails, the transmitter sends $A_2$ and $B_2$, and if $B_1$ and $B_2$ succeed, $A_2$ and $B_3$ are transmitted (as shown in Fig. 5.1).

The throughput of level-independent protocols is expected to be more than the level-dependent protocols, because in level-dependent protocols all levels with the same message words are retransmitted while in level-independent protocols, only erroneous upper-levels are retransmitted and lower-levels are used to transmit new codewords.

For NC level-independent (NC-I) HARQ, the LLRs are calculated according to (5.2), the same as for NC-D. For CC level-independent (CC-I) HARQ, the combining scheme is a little different than that of CC-D, because different codewords may be transmitted on each level during each transmission, so the LLRs are given by

$$\lambda_{b,i,l} = \ln \frac{\sum_{s \in \mathcal{X}_b^0} \mathrm{P}(y_{i,l} \mid s_{i,l} = s)}{\sum_{s \in \mathcal{X}_b^1} \mathrm{P}(y_{i,l} \mid s_{i,l} = s)} + \lambda_{b,i,l-1}. \tag{5.4}$$

For all protocols, the system uses AMC employing modulations and codes of different spectral efficiencies and rates for each SNR. The AMC is applied through the CSIT-based mode-selection at the transmitter to enhance the throughput.

## 5.3 Throughput as a Design Metric

Typically, for the AWGN channel, polar codes are designed to minimize the FER for a given code rate, $R = \frac{K}{N}$, at a given SNR. That is, the $K$ elements of the information set are chosen in an attempt to provide as low a FER as possible. Alternatively, for a given SNR and a target FER, one can choose the information set to be as large as possible (thereby maximizing the code rate), while ensuring the target FER is not exceeded. However, for systems employing HARQ, neither the FER nor the code rate is of primary importance. For ARQ systems, messages are transmitted indefinitely until they are correctly received. As such, the throughput, which is the rate that message bits are correctly received (in message bits per channel use), is a more relevant metric. By using this new criterion, codes that are quite different from those that maximize the code rate or minimize the FER can be designed.

For the NC-D protocol, where a single code is used across all binary channels of the constellation and received samples from previous transmission are not exploited, the throughput is given by

$$\eta_{\text{NC-D}} = \frac{K}{NB}(1 - P_K), \tag{5.5}$$

where $P_K$ is the FER when the message word length is $K$. The optimization problem, for a given codeword length, $N_{tot} = NB$, is to find $K$, and the associated information set of the polar code, that maximize $\eta_{NC-D}$. When the NC-I protocol is used, the throughput, since the levels are independent, is

$$\eta_{\text{NC-I}} = \sum_{b=1}^{B} \frac{K_b}{N}(1 - P_b^{K_b}), \tag{5.6}$$

where $K_b$ is the length of the information set of the codeword transmitted on the $n^{th}$ binary channel of the constellation (with $K = \sum_b K_b$ still being the total number of transmitted message bits.) and $P_b^{K_b}$ is the corresponding FER of that code.

When Chase-combining is used, the FER decreases depending on the number of retransmissions, so the throughput of the CC-D protocol is

$$\eta_{\text{CC-D}} = \frac{K}{NB\sum_{l=1}^{L}\prod_{l'=1}^{l-1}P_{l'}^{K}}(1-P_{L}^{K}), \tag{5.7}$$

where $P_l^K$ is the FER of the $l^{th}$ transmission, given that the previous $l-1$ transmissions of the codeword failed. For CC-I, the throughput is

$$\eta_{\text{CC-I}} = \sum_{b=1}^{B}\frac{K_b}{N\sum_{l=1}^{L}\prod_{l'=1}^{l-1}P_{b,l'}^{K_b}}(1-P_{b,L}^{K_b}), \tag{5.8}$$

where $P_{b,l}^{K_b}$ is the FER of the $l^{th}$ transmission at $b^{th}$ level, given that the previous $l-1$ transmissions of the codeword failed.

## 5.4 Polar Code Design Methods for BPSK

In this section, we start by explaining the simulation-based design method for SCD. Then, we describe the design method for NC and CC HARQ using GA which can design the code with low-complexity.

### 5.4.1 Simulation-based Code Design for SCD

Using the method described in Section 4.3.1, it is straightforward to design polar codes to maximize the throughput. Once the simulation of a sufficiently large number of codewords has completed (typically $N_{\text{SIM}} = 10000$ codewords is sufficient for low-to-moderate SNRs) at the desired SNR and the position of the first error events has been recorded, the information set of the minimum FER polar codes for every code rate from $\frac{1}{N}$ to 1 is determined (i.e., $\forall K \in \{1, ..., N\}$) and the corresponding

FER is approximated. Then the code rate that maximizes the throughput, (5.5), is determined, and the associated information set is used to define the optimal polar code at that SNR.

### 5.4.2 GA-based Code Design for SCD

Here, we design the throughput-optimal polar codes using the GA and compare it with the simulation-based approach. For designing the code we call Function NC-Binary-Design(GA-BER($4\times10^{\gamma/10}$,$N$),$N$). Function NC-Binary-Design approximates the FER of the code based on the bound given in (4.2). In the next step, the corresponding throughput for each message-length is computed and the code rate with the highest throughput is chosen. From (5.5) and (4.2), the throughput can be written as

$$\eta_{\mathrm{NC},K} = \frac{K}{NB}\prod_{i=1}^{K}(1-v_i). \tag{5.9}$$

**Theorem 3.** *The FER in (4.2) is monotonically increasing with respect to $K$.*

*Proof.* The FER estimation in (4.2) can be written in a recursive form as

$$P_K = 1-(1-v_K)\prod_{i=1}^{K-1}(1-v_i) = P_{K-1} + v_K\prod_{i=1}^{K-1}(1-v_i). \tag{5.10}$$

Therefore, (5.10) monotonically increases with $K$ since $v_K\prod_{i=1}^{K-1}(1-v_i) > 0$ due to $0 < v_i < 1$. ∎

**Theorem 4.** *The throughput in (5.5) is a unimodal function of the code rate.*

*Proof.* The first forward difference $\Delta_K$ can be given as

$$\Delta_K\eta = \eta_{K+1} - \eta_K = \frac{1}{NB}\big[1-(K+1)v_{K+1}\big]\prod_{i=1}^{K}(1-v_i). \tag{5.11}$$

Since $\Delta\eta_K > 0$ for $K < \frac{1}{v_{K+1}} - 1$ and $\Delta\eta_K < 0$ for $K > \frac{1}{v_{K+1}} - 1$, the difference equation has only one root, so $\eta_K$ has only one maxima. Thus, (5.5) is unimodal. ■

**Function** NC-Binary-Design($\mathbf{v}$,$N$)

**Input:** $\mathbf{v}$ and $N$

**Output:** The sorted bit-channel indices **idx** and the code rate $R_{\text{SCD}}$

1 [$\mathbf{v}$,**idx**]=Sort($\mathbf{v}$) ▷ Sorting the bit-channels, defined in Algorithm 2.

2 **for** $\kappa = 1 : N$ **do**

3 $\quad P_\kappa = 1 - \prod_{i=1}^{\kappa}(1 - v_i)$ ▷ FER estimation for all Ks

4 $\quad \eta_\kappa = (1 - P_\kappa)\frac{\kappa}{N}$ ▷ Estimation of $\eta_\kappa$

5 $K_{opt} = \arg\max_{\kappa} \eta_\kappa$ ▷ Finding $K$ with highest $\eta_\kappa$

6 $R = \frac{K_{opt}}{N}$

7 return **idx**, $R_{\text{SCD}}$

The throughput vs. the code rate for polar codes with length 4096 at SNRs of -2, 0 and 2 is shown in Fig. 5.2. From Theorem 3, we know $P_K > P_{K-1}$. Thus, the throughput initially grows (nearly linearly for the FER-minimal codes) with the code rate until the rate gets sufficiently high that the effects of the FER start to dominate in (5.5), after which point the throughput drops dramatically. The existence of an optimal rate to maximize the throughput is clear which is shown in Theorem 4.

The FER of the first transmission of throughput-maximizing codes is shown in Fig. 5.3. Observe that the FER of the first transmission is high at low SNRs.

### 5.4.3 Code Design for Chase-combining

To design the code for the CC HARQ schemes, the code design method for NC should be modified since the average LLR with each retransmission increases to $\frac{4l}{N_0}$, where $l \in \mathbb{N}$ is the retransmission attempt number. In this case, the procedure of the design includes running the Function GA-BER to estimate the order of bit-channels for the

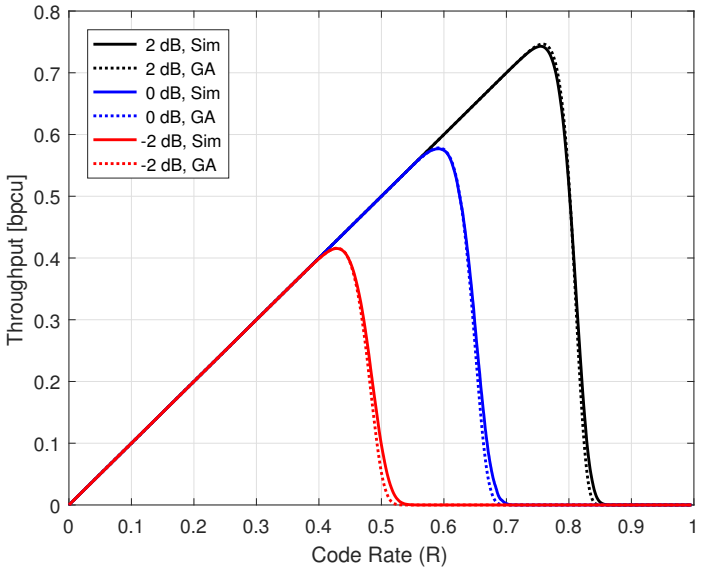



**Figure 5.2:** Throughput of NC protocols vs. the code rate with $N = 4096$ bits at different SNRs for BPSK with SCD.

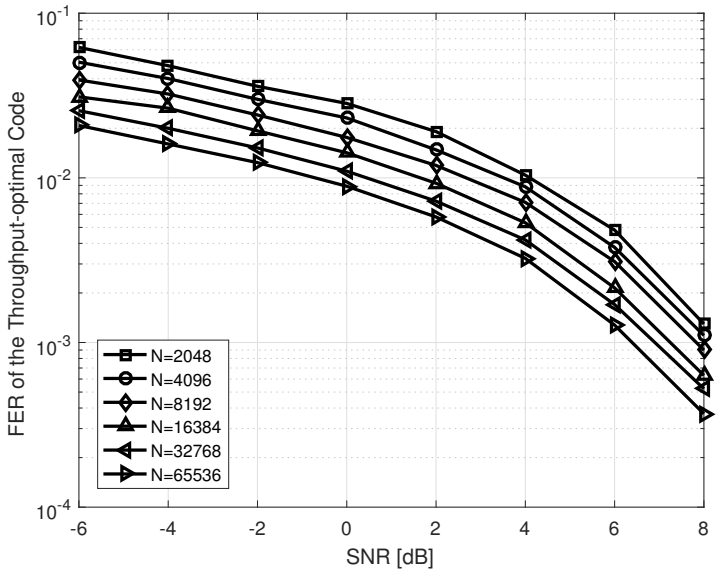



**Figure 5.3:** Comparison of the FERs of the first transmission of throughput-optimal codes for NC protocol with BPSK.

first transmission since the majority of codewords are decoded correctly in the first step. To find the rate corresponding to the maximum throughput, the FERs using the Function GA-BER for each retransmission are estimated and (5.7) is computed. The saturation of the maximum throughput is used as the stopping criterion. The code design for CC HARQ is described formally in Function CC-Binary-Design. To design the code, we should call Function CC-Binary-Design($\gamma$,$N$). The codes designed for CC have slightly higher rates than the codes designed for NC. The throughput of the CC and NC are compared in Fig. 5.4.

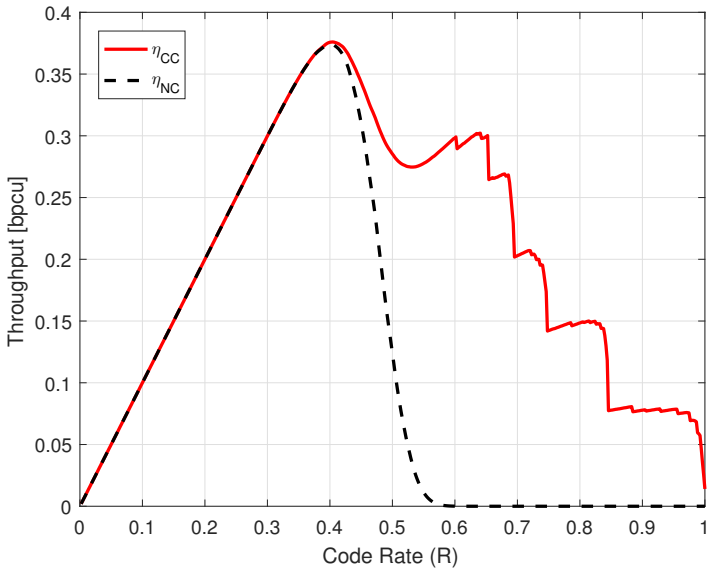



**Figure 5.4:** Throughput of NC and CC HARQ vs. the code rate with $N = 512$ bits at $\gamma = -2$ dB for a BPSK and a SCD.

## 5.5 Polar Code Design Methods for QAM

To design MLPCM for PAM and QAM constellations, the GA method can be employed. In this section, we explain the design steps of MLPCM and provide algorithms to maximize the throughput of HARQ schemes.

**Function** CC-Binary-Design($\gamma$,$N$)

**Input :** $\gamma$ and $N$

**Output:** The maximum throughput code for the CC HARQ and the corresponding code rate for the SCD

$\eta' = 1$, $\eta = 0$, $\bar{N}_0^\kappa = 0$, $P_0^\kappa = 0$,$l = 1$

**while** $\eta' \neq \eta$ **do**

  $\bar{\lambda} = 4l \times 10^{\gamma/10}$

  $\mathbf{v}$=GA-BER($\bar{\lambda}$,$N$) ▷ Determining the conditional BERs

  **if** $l = 1$ **then**

    [$\mathbf{v}$,**idx**]=Sort($\mathbf{v}$) ▷ Sorting the bit-channels

  **else**

    $\mathbf{v} = \mathbf{v}(\mathbf{idx})$ ▷ Sorting the bit-channels based on the **idx** of the first transmission

  **for** $\kappa = 1 : N$ **do**

    $P_l^\kappa = 1 - \prod_{i=1}^{\kappa}(1 - v_i)$ ▷ FER estimation

    $\bar{N}_l^\kappa = \bar{N}_{l-1}^\kappa + N \prod_{l'=1}^{l-1} P_{l'}^\kappa$ ▷ Vector of effective code lengths

    $\bar{\eta}_\kappa = \frac{\kappa}{\bar{N}_l^\kappa}(1 - P_l^\kappa)$ ▷ Vector of throughputs for all message lengths

  $K_{opt} = \arg\max_\kappa \bar{\eta}_\kappa$ ▷ Finding $K$ with the highest $\eta_\kappa$

  $\eta' = \eta$

  $\eta = \bar{\eta}_{K_{opt}}$ ▷ Finding the highest throughput

  $l = l + 1$ ▷ Incrementing retransmission number

$R_{\text{SCD}} = \frac{K_{opt}}{N}$

return **idx**, $R_{\text{SCD}}$

### 5.5.1 GA-based Code Design for SCD

To construct polar codes using GA, the average LLR for each binary channel of a QAM constellation is estimated using the method in Section 4.3.2. The bound (5.9) can be extended for MLPCM to

$$\eta_{\text{NC}} = \frac{K}{N} \prod_{b=1}^{B} \prod_{i=1}^{K_b} (1 - v_b^i). \tag{5.12}$$

From (5.12), to determine component code rates for MLPCM, the bit-channel reliabilities $v_b^i, \forall i = 1, ..., N$ can be measured for all levels $b = 1, ..., B$ and among them, those with the lowest genie-aided BERs are chosen as the total information set. Thus, the component code rates are automatically determined. The entire code design procedure for the NC-D protocol is mentioned in Algorithm 6.

**Algorithm 6:** GA-based design for NC-D with QAM

**Input** : $\bar{\boldsymbol{\lambda}}_{PAM}$: The average LLR vector for PAM, R, B and N
**Output:** The maximum throughput codes and the code rate for the SCD
$\bar{\boldsymbol{\lambda}}_{\boldsymbol{QAM}}$=AverageLLR($\bar{\boldsymbol{\lambda}}_{PAM}$) ▷ Defined in Section 4.3.2.
**for** $b = 1 : B$ **do**
  $\mathbf{v}_b$=GA-BER($\bar{\lambda}_{QAM,b}$,$N$) ▷ Determining conditional BERs of each level
[$\mathbf{idx}_{tot}$,$R$]=NC-Binary-Design([$\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_B$],$BN$) ▷ Defined in Section 5.4.2.
return $\mathbf{idx}_{tot}$, $R_{\text{SCD}}$

An algorithm for designing throughput optimal codes for NC-I protocol is presented in Algorithm 7. Due to the independence of levels, the algorithm designs a binary code for each level independently. The same algorithm can be developed for the CC-I protocol by substituting the Function CC-Binary-Design instead of the Function NC-Binary-Design in Algorithm 7.

**Algorithm 7:** GA-based design for NC-I with QAM

**Input** : $\bar{\boldsymbol{\lambda}}_{\text{PAM}}$: The average LLR vector for PAM and $N$
**Output:** The maximum throughput codes and the code rates for the SCD
$\bar{\boldsymbol{\lambda}}_{\text{QAM}}$=AverageLLR($\bar{\boldsymbol{\lambda}}_{\text{PAM}}$) ▷ Defined in Section 4.3.2.
**for** $b = 1 : B$ **do**
  $\mathbf{v}_b$=GA-BER($\bar{\lambda}_{\text{QAM},b}$,$N$) ▷ Defined in Section 4.3.1.
  [$\mathbf{idx}_b$,$R_{\text{SCD,b}}$]=NC-Binary-Design($\mathbf{v}_b$,$N$) ▷ Defined in Section 5.4.2.
return $\mathbf{idx}_1$,..., $\mathbf{idx}_B$, $\mathbf{R}_{\text{SCD}}$

We do not mention the design algorithm for the CC-D protocol but it is only the extension of Function CC-Binary-Design and Algorithm 6.

## 5.6 Rate Matching Algorithm for SCLD

For decoding of each output bit with SCD, the information of other previously decoded bits and the future information bits are not used. To overcome these shortcomings, the SCLD records a list containing different possible decoded message words and keeps only $L$ most likely ones after each step [62]. A CRC sequence is usually added to message bits when SCLD is used, to increase the probability of finding the most likely message word. Throughout this chapter for SCLD, only one CRC sequence is used for both list decoding and ARQ. Typically, the codes designed for SCD are used for SCLD as well since the SCL core decoder is SCD. However, these codes are suboptimal for SCLD.

When throughput-maximizing codes optimized for SCD are used with SCLD, the FER is lower than the FER of SCD. Even though this slightly improves the throughput, it is not highly effective on the term $(1-P_K)$ in (5.5). However, since $R$ is numerically more dominant in (5.5) when the FER is small, it can be increased more significantly to improve the throughput. Therefore, we introduce a rate matching algorithm for SCLD. The algorithm employs the golden-section search method to find the code rate corresponding to the maximum throughput.

The golden-section search method iteratively measures the objective function at different points and updates the answer range interval $[a, b]$ until this interval is narrowed down around the final value of the decision variable. For a detailed explanation of the golden-section search method refer to [122] and references therein. Here, the objective function is the actual throughput of SCLD measured using simulation and the decision variable is the message word length. The proposed algorithm is fast, e.g., for $N = 16384$ it finds the optimum rate in around 15 iterations, corresponding to 16 evaluations of the objective function. For initialization of the algorithm, we use $a = K_{\text{SCD}}$, the optimal message word length for SCD, and

$b = \min(a + \frac{BN}{10}, BN)$. The algorithm is formally presented in Function SCLD-Rate-Match. Note than when we use the algorithm for the NC-D protocol, we should call Function SCLD-Rate-Match($\gamma$,$R$,$N$,$\mathbf{idx}_{tot}$,$B$) to repeat the simulation for all $B$ binary channels of a constellation. However, for the NC-I protocol we call SCLD-Rate-Match($10\log_{10}\frac{\bar{\lambda}_b}{4}$,$R$,$N$,$\mathbf{idx}_b$,1) for each binary channel $b = 1, ..., B$, independently.

A comparison of throughput vs. the code rate for SCD and SCLD at different lengths is presented in Fig. 5.5. Observe that the rate-matched codes for SCLD substantially improve the throughput for all code lengths. Furthermore, as we increase $N$, the throughput tends to the capacity at which the throughput and the code rate of the throughput-maximizing code eventually are equal.

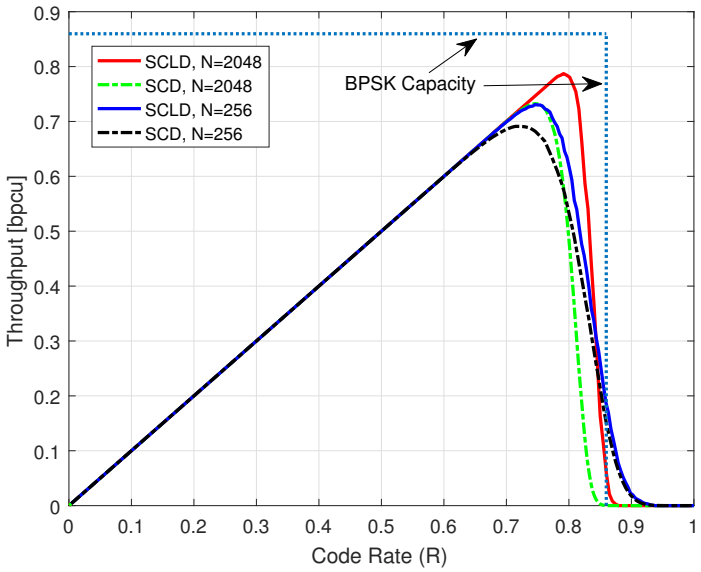



**Figure 5.5:** Throughput of NC protocol vs. the code rate with different code lengths at $\gamma = 2$ dB for BPSK.

**Function** SCLD-Rate-Match($\gamma$,$R$,$N$,**idx**,$B$)

**Input :** $\gamma$, $R_{SCD}$, $N$, **idx**, and $B$
**Output:** The code rate of the maximum throughput code for SCLD
**Procedures:**
f($K$)=SCLD_Throughput($\gamma$,$K$,$N$,**idx**,$B$): Employs SCLD to simulate and estimate the throughput of a MLPCM with a total message length of $K$, a level-code length of $N$, the indices **idx** designed for SCD, a $2^B$-QAM constellation, and SNR $\gamma$.
Golden ratio: $\rho = \frac{\sqrt{5}-1}{2}$

$K_{\text{SCD}} = \lfloor R_{SCD} BN \rfloor$
$a = K_{\text{SCD}}$ ▷ Initialization
$b = \min(a + \frac{BN}{10}, BN)$ ▷ Initialization
$k_1 = \lfloor \rho a + (1-\rho) b \rfloor$ ▷ The first point to test in the interval
$f(k_1)$ =SCLD_Throughput($\gamma$,$k_1$,$N$,**idx**,$B$) ▷ The throughput for K=k1
$k_2 = \lfloor (1-\rho) a + \rho b \rfloor$ ▷ The second point to test in the interval
$f(k_2)$ = SCLD_Throughput($\gamma$,$k_2$,$N$,**idx**,$B$) ▷ The throughput for K=k2
**while** $|a-b|^2 > 1$ **do**
  **if** $f(k_1) > f(k_2)$ **then**
    $b = k_2,\ k_2 = k_1,\ f(k_2) = f(k_1)$
    $k_1 = \lfloor \rho a + (1-\rho) b \rfloor$ ▷ Updating the first message length
    $f(k_1)$ = SCLD_Throughput($\gamma$,$k_1$,$N$,**idx**,$B$)
  **else**
    $a = k_1,\ k_1 = k_2,\ f(k_1) = f(k_2)$
    $k_2 = \lfloor (1-\rho) a + \rho b \rfloor$ ▷ Updating the second message length
    $f(k_2)$ = SCLD_Throughput($\gamma$,$k_2$,$N$,**idx**,$B$)
$R_{SCLD} = \frac{k_1}{BN}$
return $R_{SCLD}$

## 5.7 MLPCM Design for MIMO-SVD

MIMO-SVD [119, 123, 124] is a category of closed-loop MIMO schemes that uses a feedback to achieve CSIT. To apply this scheme, assuming that $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathcal{H}}$, the SVD of a realization of $\mathbf{H}$ is taken to find matrices $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$. Then, the vector of modulated symbols is multiplied by $\mathbf{V}$ and transmitted over the channel. At the receiver, $\mathbf{U}^{\mathcal{H}}$ is multiplied to the received signal and the resulted signal is given as

$$\mathbf{Y} = \mathbf{U}^{\mathcal{H}}\big(\mathbf{HVs} + \mathbf{w}\big) = \mathbf{\Sigma}\boldsymbol{s} + \tilde{\mathbf{w}}, \tag{5.13}$$

where $\tilde{\mathbf{w}} = \mathbf{U}^{\mathcal{H}}\mathbf{w}$. Note that $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{w}}$ have the same statistical properties [119]. Using the SVD, the channel $\mathbf{H}$ is decomposed into independent parallel channels known as eigenmodes of the channel. The optimal power allocation for MIMO-SVD is the water-filling approach [124]. Assuming $\mathbf{H}$ is full rank and $N_t \leq N_r$, the number of eigenmodes is the rank of $\mathbf{H}$ which is $N_t$. Thus, $\mathbf{s}$ can be given as

$$\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \dots & s_{N_t} \end{bmatrix}, \tag{5.14}$$

where $s_i$ is a symbol chosen from a QAM. When MIMO-SVD is used, the codes designed for the throughput in Section 5.5 can be employed based on the SNR of each parallel channel. In this case, at each average SNR in fading channels, achieving the ergodic MIMO capacity is possible as the code length tends to infinity. Thus, the average throughput can be used as a performance measure for slow fading channels with full CSIT. The average throughput for AMC is given by

$$\bar{\eta} = \frac{1}{N_F} \sum_{n_f=1}^{N_F} \sum_{i=1}^{N_t} \frac{K_i(\bar{\gamma}, \mathbf{H}_{n_f})}{N B_i(\bar{\gamma}, \mathbf{H}_{n_f})} \Big(1 - P_{K_i,i}(\bar{\gamma}, \mathbf{H}_{n_f})\Big) \tag{5.15}$$

## 5.8 Numerical Results and Discussions

In this section, we provide the performance analysis of the code design algorithms described in Sections 5.4 and 5.5, respectively. The system described in Section 5.1 is used for all simulations and the CRC sequence is CRC-16-CCITT. Note that all curves for AWGN and MIMO fading channels are sketched based on the transmitted instantaneous SNR and the transmitted average SNR, respectively.

In Fig. 5.6, the throughput of SC-decoded polar codes with different lengths changing from 4096 bits to 1048576 bits are shown. At 0 dB, the polar codes of lengths 1048576 and 4096 achieve 94.8% and 80% of the capacity, respectively.

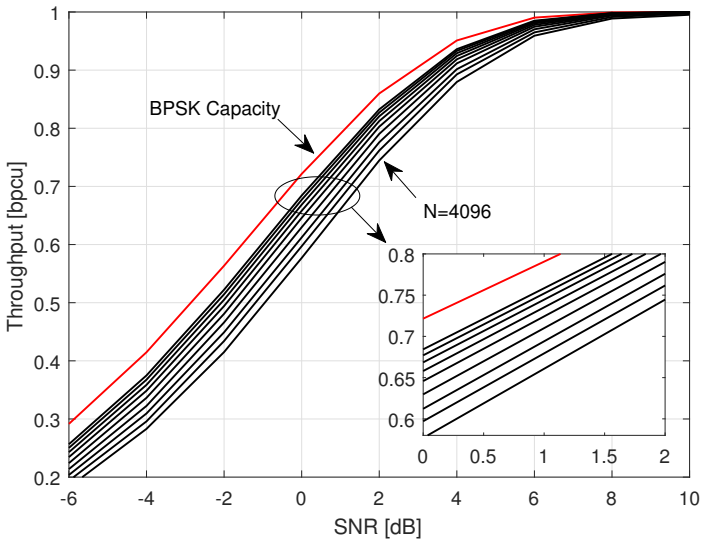



**Figure 5.6:** Throughput comparison of polar codes of different lengths with SCD.

The throughput of SC- and SCL-decoded polar codes of length 4096 is shown in Fig. 5.7 in comparison to BPSK capacity. The SCLD list size is 32 for all curves. The lowermost black curve shows the throughput of the polar code designed using SCD and decoded with SCD that achieves the throughput of 80% of the capacity at 0 dB. The second black curve is the throughput of the code designed for SCD and decoded using SCLD which achieves 82.5% of the capacity at 0 dB. The topmost curve under the capacity shows the performance of the code designed for SCD and rate matched for SCLD which achieves the 89.3% of the capacity at 0 dB. Therefore, the use of SCLD for decoding of codes designed for SCD does not change the throughput substantially in comparison to SCD. However, employing Function SCLD-Rate-Match for the rate matching can substantially improve the throughput of the code used with SCLD.

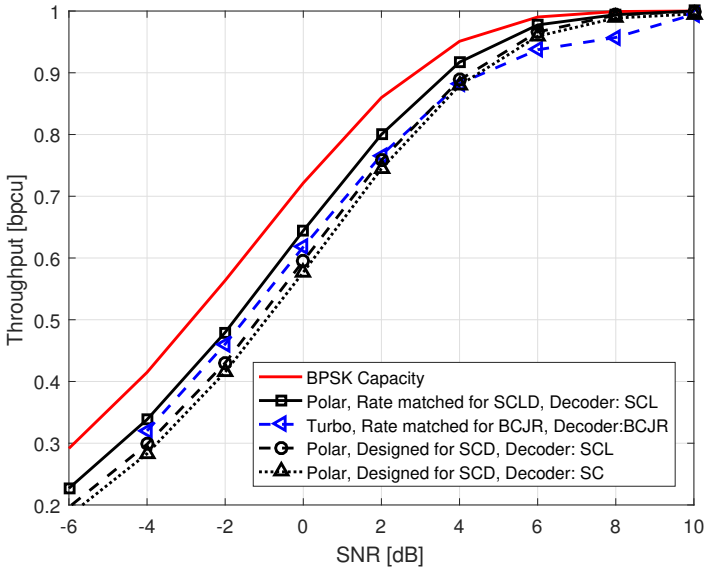



**Figure 5.7:** Throughput comparison of polar codes of length 4096 optimized for SCD decoded with SCD and SCLD, and the SCLD rate matched codes decoded with SCLD.

Fig. 5.7 furthermore provides a comparison of the throughput-maximizing polar codes and parallel concatenated (turbo) codes employed in LTE-A [14]. The BCJR decoder with 5 iterations is employed for decoding of the turbo codes with codeword lengths of around 4096. The LTE-A turbo code rate is optimized using the golden-section search (similar to Function SCLD-Rate-Match but the BCJR is employed) to maximize the throughput. The range of message word lengths is limited to 40:8:512, 528:16:1024, 1056:32:2048 and 2112:64:4200 bits which provides us with 157 different choices for the code rate. In this case, a golden-section search is used to search all the possible 157 choices for the code rate and selects the code rate corresponding to the highest throughput. Due to code rate limitations, the optimization procedure was only applied in the SNR range between -4 and 10 dB. Furthermore, the turbo code lengths are slightly higher than 4096. It can be observed that turbo code performance is close to the optimized polar code using NC-Binary-Design at low SNRs. However,

as the SNR increases, the performance of turbo degrades and at high SNRs, it is even worse than the polar code optimized for SCD. Note that the complexity of BCJR with 5 iterations is more than that of SCLD.

In Fig. 5.8, the code rate $R_{\text{SCLD}}$ of the rate matched codes designed using Function SCLD-Rate-Match plotted against the code rate of the code optimized for SCD for a variety of code lengths. Observe that the change of $R_{\text{SCLD}}$ against $R_{\text{SCD}}$ is approximately linear for all code lengths.

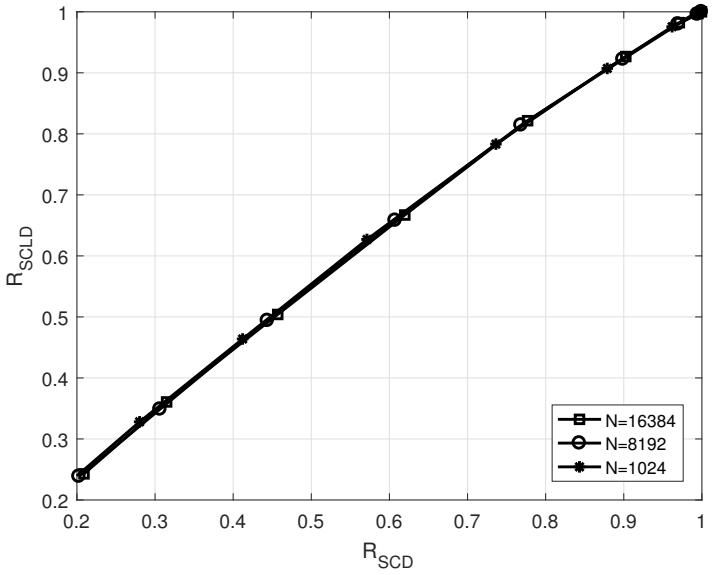



**Figure 5.8:** Comparison of $R_{\text{SCLD}}$ found using Function SCLD-Rate-Match against $R_{\text{SCD}}$ for a variety of $N$.

The comparison of the CC HARQ schemes proposed in [118] for SCD and [117] for SCLD and the CC HARQ design method introduced in Function CC-Binary-Design is provided in Fig. 5.9 for BPSK. To construct the polar codes using algorithms provided in [118] and [117], the $K$ at each SNR is considered the same as $K$ found using Function CC-Binary-Design. The results indicate that the methods proposed in [118] and [117] cannot generate codes better than the Function CC-Binary-Design. In addition, the NC codes optimized using Function NC-Binary-Design have approximately

the same performance as the codes generated for the CC HARQ using Function CC-Binary-Design for situations where we can adapt the code when the SNR changes. In this case, CC does not have any sensible advantage over NC since when we design the codes for NC or CC, the algorithms try to minimize the number of retransmissions and keep it in order of at most one to maximize the throughput.

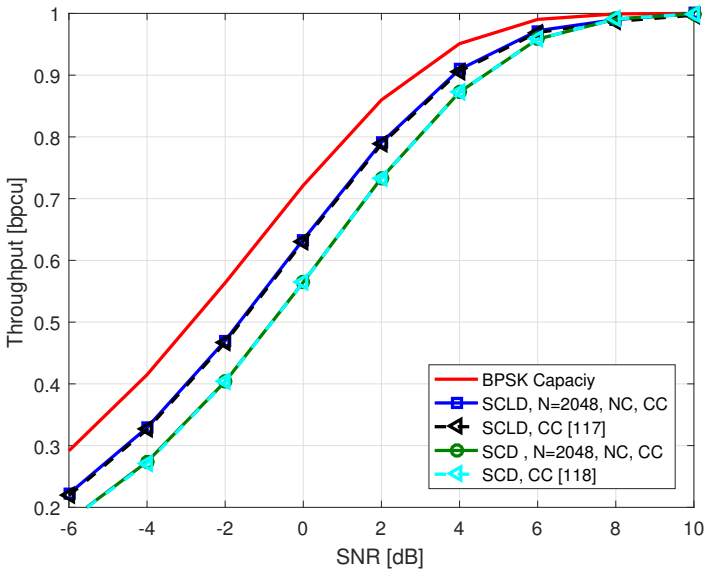



**Figure 5.9:** Comparison of proposed NC and CC codes and code design methods in [118] and [117] for SCD and SCLD, respectively.

The performance of MLPCM designed for NC-D protocol using Algorithm 6 with a SCD is shown in Fig. 5.10 for BPSK, 4-QAM, 16-QAM, and 64-QAM constellations. For all SNRs, the throughput of the code is essentially identical regardless of whether the exact LLR calculation or the simplified LLR calculation of Section 3.1.1 is used. This is despite the fact that the simplified LLR calculation is not exact at low SNRs.

The comparison of NC-I and NC-D protocols with SCD and SCLD is presented in Fig. 5.11 for 16-QAM. At all SNRs, NC-I and NC-D with SCLD perform slightly better than the corresponding protocol with SCD. At 4 dB, NC-D with SCD and SCLD achieves 69% and 73% of the capacity and NC-I with SCD and SCLD achieves

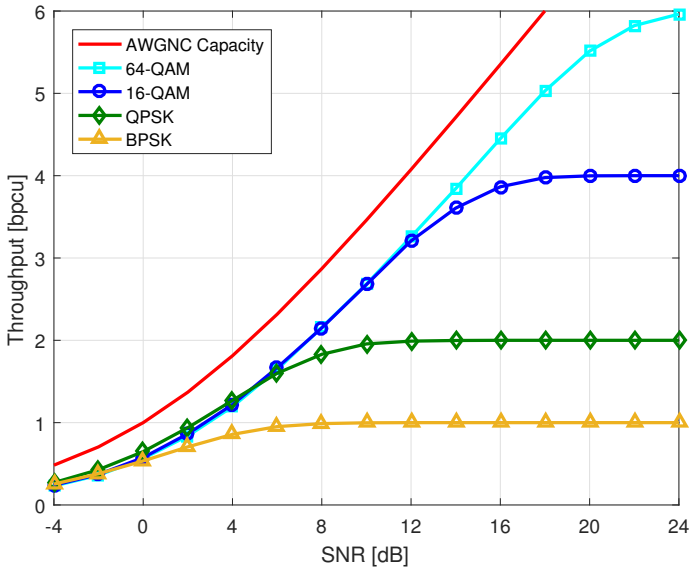



**Figure 5.10:** Throughput of MLPCM designed using Algorithm 3 with $N = 512$ for BPSK and a variety of QAM constellations.

74% and 84% of the capacity, respectively.

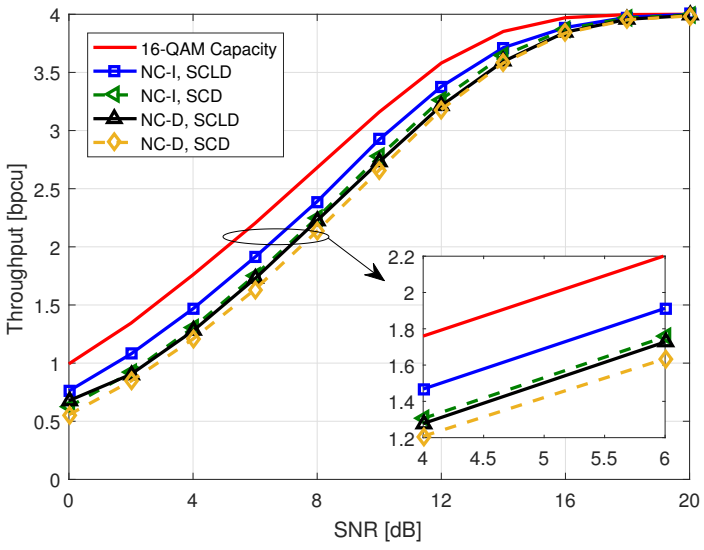



**Figure 5.11:** Throughput comparison of NC-D and NC-I protocols with SCD and SCLD for $N_{tot} = 2048$ and 16-QAM.

In all previous results, we assumed we can adaptively change the code and the modulation. In those cases, CC HARQ protocols do not provide any advantage over NC protocols. However, once the number of codes is limited, the advantage of CC HARQ is highlighted. In Fig. 5.12, the performance of CC-I protocol with two codes designed at 4 dB and 14 dB is compared with the IR HARQ and CC HARQ schemes constructed using polar codes and BICM in [120] for 16-QAM. We observe that when we use the code designed at 14 dB for the whole range of SNRs, IR HARQ is better than the proposed throughput-optimal codes for CC-I protocol at a few SNRs. However, on average the proposed MLPCM with CC-I protocol performs up to 3 dB better than IR HARQ and CC HARQ in [120]. In case we use one more code designed at 4 dB, CC-I at all SNRs achieves higher throughput than schemes proposed in [120]. The reason for this superiority is the good design of MLPCM scheme in conjugation with CC-I scheme.

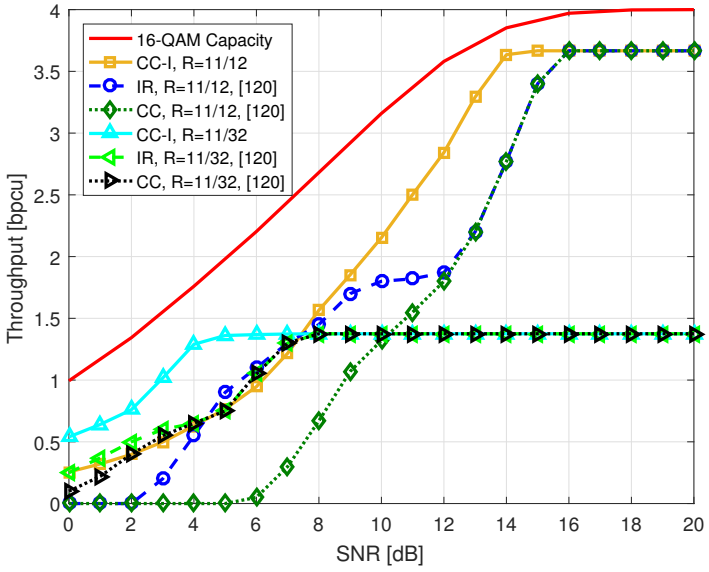



**Figure 5.12:** Comparison of CC-I with two codes and the IR and CC HARQ proposed in [120].

Fig. 5.13 shows the average throughput of the MIMO-SVD with AMC and $N_t$ =

$N_r = 2$. For AMC, we employed QAMs with cardinalities 4, 16, and 256. For BIPCM, $N_{tot}$ is set to 2048 bits and for MLPCM, $N$ is set to 128 bits. Thus, the MLPCM code length is variable and its maximum is 2048 bits. Results show that the MLPCM scheme works within 1.5 dB and 2 dB of the scheme's capacity at low and high SNRs, respectively. Furthermore, MLPCM outperforms BIPCM for up to 10% of the scheme's capacity. Note that at low SNRs, most of the time, AMC selects 4-QAM and 4-QAM performance is approximately the same for both BICM and MLC/MSD as shown in Fig. 4.3. Thus, the throughput of MLPCM and BIPCM is approximately the same at low SNRs.

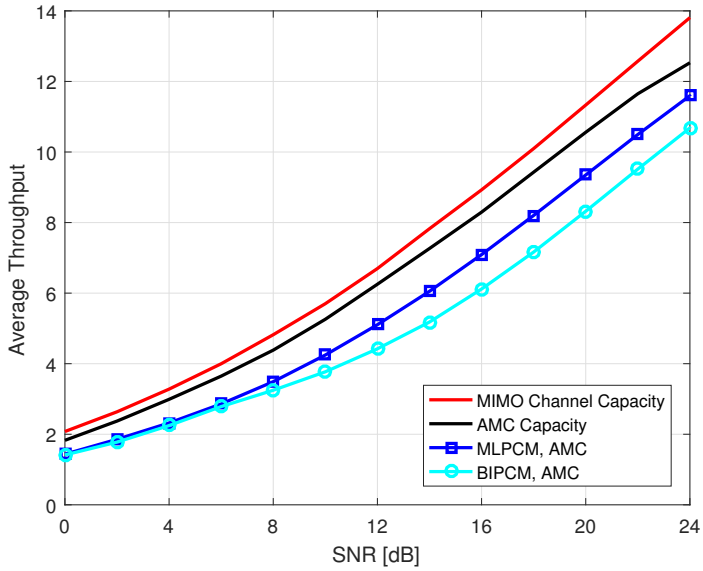



**Figure 5.13:** Average throughput for MIMO-SVD with AMC constructed using QAMs with a variety of cardinalities.

## 5.9 Conclusion

In this chapter, we proposed a set of algorithms for designing MLPCM by maximizing the throughput for NC and CC HARQ schemes. The numerical results show the codes

constructed using these methods perform very close to the capacity, e.g., within 1.2 dB of the 16-QAM capacity with $N = 2048$ bits. The results indicate an improvement up to 15% of the capacity at 4 dB when we use level-independent protocols and rate-matched SCLD. This class of polar code design can substantially enhance the performance when the CSIT is available, e.g., for MIMO-SVD with AMC.

# Chapter 6

# Uncoded Space-Time Signal Design Based on Error Bounds

In wireless communication systems, high quality, capacity and reliability are among the essential demands. One of the key enablers for this improvement is enhanced physical layer design by optimizing each block of the transceiver. Signal shape design, also known as constellations design, can substantially affect the performance of communication systems. Traditionally, 1D PAM and 2D QAM have been widely employed in the majority of wireless systems due to their simple decoding. However, optimization of signal constellations provides us with better matching between signaling and the communication channels, which may substantially improve the performance of the system. Improving the performance of a communication system by finding a better placement of the constellation points has been known as the packing problem and has been studied widely in literature by using mathematical tools such as lattice constructions, where lattice-based constellation design is used for finding the densest packing [125, 126]. Recently, Beko and Dinis in [127] revisited the problem of designing multidimensional constellations by using contemporary optimization tools, whereby they minimized the sum power of all points with a constraint on the minimum distance between points. In spite of generating good constellations, there is

still room for improvement as this method does not consider minimizing the number of neighbors of each constellation point, known as the kissing number [125].

To achieve even better performance, signal constellations can be optimized with respect to the SER or BLER expressions if such expressions are available. Since the exact error rate expressions are difficult to derive in most cases, constellation optimization based on either approximate expressions or bounds may prove to be a feasible alternative. A low-complexity class of upper bounds on the performance of orthogonal transmission schemes which assumes an arbitrary position for constellation points can be derived by using the Chernoff bound on PEP at high SNR [128]. For deriving any closed-form bound on the performance in a fading channel, knowing the distribution of the fading and the number of antennas of the transceiver are essential. As a result, constellations are only designed for a specific channel.

It is well known that the bit-to-symbol mapping plays a critical role in improving the BER. The traditional approach in the literature for finding the BER-minimizing constellations includes two sequential steps: optimization of the constellation shape followed by optimization of the bit-to-symbol mapping [129]. However, constellation shapes optimized using a bound on BER can substantially improve the BER performance [129]. Since the main source of bit errors at high SNR is due to the received sample falling within a decision region that is adjacent to the transmitted symbol, the SER or BLER bounds can be extended to a bound on the BER for high SNR values by considering the Hamming distance [9, 97]. As explained in Section 6.3, this bound can be further used for the optimization of the constellation shape.

Signal constellations in which points are mapped to more than two dimensions, hereafter called multidimensional constellations, allow for increased separation between points in comparison to the widely used 2D constellations [130]. Multidimensional constellations can be projected onto a set of orthogonal 2D signal spaces, with

each projection transmitted independently. For example, a 4D symbol can be transmitted using two 2D symbols. Designing multidimensional constellations has been discussed using different techniques: in [83, 126, 131] based on lattice construction; in [132] based on the behavior of charged electrons in free space; and in [127] based on the optimization of non-lattice construction. However, as mentioned earlier, optimization of the bounds on the error rates of these constellations can further improve the performance.

To achieve better performance with multidimensional constellations, the constellation can be designed according to the method of transmitting different dimensions of each constellation point. One way of transmitting multidimensional constellations involves using OSTBCs. OSTBCs, as one of the main types of STBCs, are used to provide full diversity with a linear complexity decoder [110,133,134]. Because of their low-complexity in encoding and decoding, OSTBCs have been used widely in standards [34]. However, the Alamouti scheme with two transmit antennas is the only full-rate full-diversity OSTBC; the other OSTBCs suffer from a rate loss in order to preserve their orthogonal structure.

During the last decade, there have been extensive studies on designing high-performance high-rate space-time codes, and some STBCs with higher performance and higher complexity, such as quasi-orthogonal STBCs (QOSTBCs) [135–137] and algebraic codes [138, references therein] have been introduced. As a major shortcoming, quasi-orthogonal and algebraic codes are designed for high spectral efficiencies and show rather poor performance for low-to-moderate spectral efficiencies. Furthermore, despite the existence of few studies such as [139] on designing STBCs with fewer receive antennas ($N_r$) than transmit antennas ($N_t$), most of algebraic codes are designed or work well under $N_t \leq N_r$, and therefore are not suitable for downlink where the number of receive antennas in the user equipment is usually limited. Fortunately, due to the possibility of optimizing multidimensional constellations for

low-to-moderate spectral efficiencies and for any $N_t$ and $N_r$, OSTBC with multidimensional constellations can outperform algebraic codes in their poor performance regions. Another shortcoming of most of algebraic codes is that they are mostly designed based on the rank and determinant criteria introduced in [33], which can improve the performance in general but do not target the improvement of BLER or BER specifically. However, since multidimensional constellations can be designed for SER, BLER or BER, OSTBC with multidimensional constellations can substantially outperform algebraic codes when the degrees of freedom of algebraic codes are not significantly higher.

Typically, an OSTBC block carries $K_S$ symbols, with independent information content carried in each symbol [133]. If we employ multidimensional constellations, each 2D component of a 2K-dimensional constellation can be carried by one of the $K_S$ different symbols of the OSTBC. This also can be seen as the generalization of a sphere packing problem [140]. In this chapter, we evaluate multidimensional constellations designed based on optimizing SER or BLER[10] bounds on the performance of OSTBCs in a Nakagami-$m$ channel. The output of the optimization problem can in general be an irregular constellation. Irregular constellations, as shown in [128] in the context of constellation rearrangement for cooperative relaying, are capable of improving the performance in comparison to regular or isometric constellations.

The main contributions of this chapter are as follows: derivation of bounds, with arbitrary constellation points, on the high-SNR orthogonal transmission SER and BER in Nakagami-$m$ channels for the single-input single-output (SISO) antenna configuration where time is the enabler for carrying different dimensions of a constellation, and for systems with a multiple-input multiple-output (MIMO) antenna configuration where OSTBC is the enabler for carrying different dimensions of a constellation.;

[10]For most parts of this chapter, a block is defined as a space-time block that consists of all dimensions of a multidimensional constellation distributed in space and time. However, for the generalized scheme, introduced in Section 6.1, a block can consist of several multidimensional constellations.

derivation of the convexity conditions of the bounds for 1D constellations.; optimizing 1D and multidimensional high-SNR SER-minimizing and BER-minimizing constellations based on the derived bounds for the Nakagami-$m$ channel.; proposing a generalized class of OSTBCs used with different size 2D and multidimensional constellations. In this chapter, we limit the optimization to multidimensional constellations for orthogonal transmission in a Nakagami-$m$ channel and low-to-moderate spectral efficiencies due to the inability of available optimizers to find a good solution for problems with a large number of variables. Note that methods in this chapter can be used only for offline optimization of small MIMO systems.

In particular, we demonstrate the performance advantage of the optimized 1D, 2D and multidimensional constellations in comparison to the best known constellations in the literature, and we show how much gain is achieved by adapting the constellation in the Nakagami-$m$ channel based on the channel parameter $m$. In addition, we show that the optimization problems for the case of 1D constellations are convex under a specific condition and we explain a set of methods to solve the convex and non-convex optimization problems efficiently. Furthermore, we show that the space-time constellations optimized using the proposed bounds outperform the best known space-time constellations in a Nakagami-$m$ channel.

Constellations optimized using high-SNR error bounds benefit from high mutual information and low outage probability similar to the rank and determinant criteria since the union bound at high SNR can well model the diversity of the system. Therefore, designing MLPCM using these constellations can potentially improve the FER compared to well-known QAM and HEX constellations. Thus, in this chapter, we evaluate the performance of optimized constellations in the presence of polar codes.

The rest of the chapter is organized as follows: The system model is described in Section 6.1, union bounds on the probability of error are derived in Section 6.2, the optimization criteria and algorithms are provided in Section 6.3, simulation results

are reported in Section 6.4, and conclusions are presented in Sections 6.5.

Throughout this chapter, to uniquely identify the constellations, the format $\hat{M}$-$\hat{N}$D is used where $\hat{M}$ is the number of points and $\hat{N}$ is the number of dimensions of a constellation; to show the 2D QAM and HEX constellations, the format $\hat{M}$-QAM or $\hat{M}$-HEX is employed. For example, 16-2D represents a 2D constellation with 16 points.

## 6.1 System Model

The system considered in this chapter consists of multiple transmit antennas that use STBCs and follows the definition in Chapter 2. The average power of the transmitted matrix $\mathbf{G}$ is set to one. The Nakagami-$m$ fading distribution is used as the general model for fading statistics because it provides a good match to a wide set of empirical measurements. The corresponding SNR distribution can be expressed as

$$f_{m,\bar{\gamma}_{n_t,n_r}}\left(\gamma_{n_t,n_r}\right) = \frac{m^m}{\bar{\gamma}^m_{n_t,n_r}\Gamma(m)}\gamma^{m-1}e^{-m\gamma_{n_t,n_r}/\bar{\gamma}_{n_t,n_r}}, \tag{6.1}$$

where $\bar{\gamma}_{n_t,n_r} = E[|h_{n_t,n_r}|^2]/N_0$ is the average SNR of each path and $m$ is the shape parameter which is fixed for all paths. For simplicity, we set $E[|h_{n_t,n_r}|^2] = 1$. The Rayleigh channel, as a special case of the Nakagami-$m$ model, can be obtained by setting $m = 1$. By denoting $g^t_{n_t}$ as the space-time code symbol transmitted in time slot $t$ from antenna $n_t$, the general ML decoding rule in the receiver for the transmission of codeword $g^1_1 g^1_2 ... g^1_{N_t} ... g^T_1 g^L_2 ... g^T_{N_t}$ in an $T \times N_t$ space-time block using perfect channel state information can be expressed as the minimization of the following metric over all constellation points:

$$\sum_{t=1}^{T}\sum_{n_r=1}^{N_r}|y^t_{n_r} - \sum_{n_t=1}^{N_t} h_{n_t,n_r}g^t_{n_t}|^2. \tag{6.2}$$

where $y_{n_r}^t$ is the received sample on the $n_r^{\text{th}}$ antenna in time slot $t$. For orthogonal transmission using the space and time resources, different antenna configurations can be used such as SISO and MIMO. In a SISO configuration, the consecutive time slots may be employed as the time resources, while in MIMO, STBCs can be employed to use both space and time resources for transmission of different dimensions of multidimensional constellations.

OSTBCs are general structures that can be employed for carrying data orthogonally over fading channels. Their simplest form, proposed by Alamouti [110] for two transmit antennas, can be written as

$$\mathbf{G}_1 = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}. \tag{6.3}$$

In Code $\mathbf{G}_1$, data are mapped separately to each constellation point and carried by symbols $s_1$ and $s_2$, both of which are independent elements of a 2D constellation, $\mathcal{S}_2$. To transmit multidimensional constellations using OSTBCs, their 2D components are distributed on OSTBC symbols. By considering $s_1$ and $s_2$ used in $\mathbf{G}_1$ as carriers of the 2D components of a multidimensional constellation, Alamouti's scheme can be rewritten as

$$\mathbf{G}_2 = \begin{bmatrix} s^{(1)} & s^{(2)} \\ -s^{(2)*} & s^{(1)*} \end{bmatrix}, \tag{6.4}$$

where $s^{(\kappa)}$ is the $\kappa^{\text{th}}$ 2D component for transmission of a multidimensional symbol $\mathbf{s} = [s^{(1)}, s^{(2)}, ..., s^{(K_S)}]$ with $\mathbf{s} \in \mathcal{S}_{2K_S}$, a $2K_S$-dimensional constellation. In $\mathbf{G}_2$, data are mapped to two 2D subpoints of a 4D point and the subpoints are carried by $s^{(1)}$ and $s^{(2)}$. As an example, to provide a spectral efficiency of 2 bits-per-channel-use (bpcu), a 4-QAM constellation should be used for $s_1$ and $s_2$ in $\mathbf{G}_1$, whereas a 16-4D

constellation should be used for $\mathbf{s} = [s^{(1)}, s^{(2)}]$ in $\mathbf{G}_2$.

For the case of four-antenna transmission, the well known OSTBC presented in [56] can be rewritten for multidimensional constellations as

$$\mathbf{G}_3 = \begin{bmatrix} s^{(1)} & s^{(2)} & s^{(3)} & 0 \\ -s^{(2)*} & s^{(1)*} & 0 & s^{(3)} \\ s^{(3)*} & 0 & -s^{(1)*} & s^{(2)} \\ 0 & s^{(3)*} & -s^{(2)*} & -s^{(1)} \end{bmatrix}, \tag{6.5}$$

and, by dropping the last column of $\mathbf{G}_3$, the corresponding scheme for a three-antenna transmission of a 6D constellation can be written as

$$\mathbf{G}_4 = \begin{bmatrix} s^{(1)} & s^{(2)} & s^{(3)} \\ -s^{(2)*} & s^{(1)*} & 0 \\ s^{(3)*} & 0 & -s^{(1)*} \\ 0 & s^{(3)*} & -s^{(2)*} \end{bmatrix}. \tag{6.6}$$

By using the orthogonal structure of OSTBCs, a simplified ML decoder can detect $s_k$ according to

$$\hat{s}_\kappa = \underset{\forall s \epsilon \mathcal{S}_2}{\operatorname{argmin}} |\hat{F}_\kappa - \left( \sum_{i,j} |h_{n_t,n_r}|^2 \right) s|^2, \tag{6.7}$$

where

$$\hat{F}_\kappa = \sum_{j=1}^{N_r} \sum_{t=1}^{T} \sum_{n_t=1}^{N_t} F^t_{n_t,\kappa}(r^t_{n_r} h^*_{n_t,n_r}). \tag{6.8}$$

In (6.8), $\kappa = 1, 2, ..., K_S$ shows the index of the different symbols carried by one OSTBC block and $F^l_{i,\kappa}(x)$ can be evaluated as

$$F^{l}_{n_t,\kappa}(x) = \begin{cases} x, & \text{if } c^{l}_{i} = s_{\kappa}, \\ x^{*}, & \text{if } c^{l}_{i} = s^{*}_{\kappa}, \\ -x, & \text{if } c^{l}_{i} = -s_{\kappa}, \\ -x^{*}, & \text{if } c^{l}_{i} = -s^{*}_{\kappa}, \\ 0, & \text{otherwise.} \end{cases} \tag{6.9}$$

This simplified decoder can be used for decoding the multidimensional constellations by changing (6.7) into a summation of decoding of different 2D components of the multidimensional constellations, expressed as

$$\hat{\mathbf{s}} = \underset{\forall \mathbf{s} \epsilon \mathcal{S}_{2K_S}}{\operatorname{argmin}} \sum_{\kappa=1}^{K_S} |\hat{F}_{\kappa} - \left( \sum_{n_t,n_r} |h_{n_t,n_r}|^2 \right) s^{(\kappa)}|^2. \tag{6.10}$$

Note that the term $P_{\kappa}$ should be computed only once for each $\kappa$, as this substantially decreases the complexity of decoding in comparison to the high-performance complex codes such as the perfect codes [35, 36] in which (6.2) may need to be computed for all points of a constellation. To estimate the LLR, the simplified detection probability for OSTBC is proportional to

$$\mathrm{P}(\mathbf{Y} \mid s_{\kappa}, \mathbf{H}) \propto \exp\left( \frac{|\hat{y}_{\kappa} - \left( \sum_{i,j} |h_{n_t,n_r}|^2 \right) s_{\kappa}|^2}{N_0 \left( \sum_{i,j} |h_{n_t,n_r}|^2 \right)} \right), \tag{6.11}$$

A simple proof for the decoder is mentioned in Appendix 6.8. Up to now, only transmission of one multidimensional symbol per codeword has been discussed. However, by considering the independence of 2D symbols in the OSTBC structure, a codeword can be split to carry symbols of multiple independent multidimensional constellations with different numbers of dimensions. As an example, $\mathbf{G}_4$ can be split to carry two 2D components of one 4D constellation and one 2D constellation; the new codeword can be written as

$$\mathbf{G}_5 = \begin{bmatrix} s_1^{(1)} & s_1^{(2)} & s_2^{(1)} \\ -s_1^{(2)*} & s_1^{(1)*} & 0 \\ s_2^{(1)*} & 0 & -s_1^{(1)*} \\ 0 & s_2^{(1)} & -s_1^{(2)*} \end{bmatrix}. \tag{6.12}$$

This generalized scheme can provide a performance-complexity trade-off in comparison to the base scheme described above where we used all independent 2D symbols of an OSTBC to carry dependent 2D symbols of a multidimensional constellation. In this scheme, since all 2D resources are not used, the number of dimensions of the multidimensional constellation decreases, which reduces the complexity of each search in ML decoding. To maintain the same spectral efficiency, the number of points can also be decreased, and, therefore, the number of searches in ML decoding decreases as well. Even though the complexity reduction results in performance degradation in comparison to the base scheme, the scheme still preserves considerable gain, especially when the OSTBC has a large size. As an example for achieving the spectral efficiency of 1.5 bpcu, $\mathbf{G}_4$ can be used with a 64-6D constellation, whereas $\mathbf{G}_5$ can be used with a 16-4D constellation for $[s^{(1)}, s^{(2)}]$ and a QPSK constellation for $s_2^{(1)}$.

## 6.2 Upper Bounds on The Performance

In this section, we derive three general bounds on the performance of an OSTBC with multidimensional constellations. 1D and 2D constellations and SISO antenna configuration are special cases of this bound. Although certain bounds on performance of OSTBCs exists in the literature [141], a specific bound based on the position of points is necessary to optimize the constellation. We start with a bound on the SER. Due to the orthogonal structure of the OSTBC, its PEP is given by [56]

$$\mathrm{P}(\mathbf{s}\to\hat{\mathbf{s}}\,|\,\mathbf{H}) = \mathrm{Q}\left(\sqrt{\left(\frac{\sum_{n_r=1}^{N_r}\sum_{n_t=1}^{N_t}|h_{n_t,n_r}|^2}{2N_0}\right)\sum_{\kappa=1}^{K_C}|s^{(\kappa)}-\hat{s}^{(\kappa)}|^2}\right), \tag{6.13}$$

By using the Chernoff bound on (6.13), a union bound on the SER of OSTBCs can be written as

$$P_s \le \frac{1}{2^B}\sum_{v=1}^{2^B}\sum_{\substack{v'=1\\v'\neq v}}^{2^B}\frac{(4)^{mN_tN_r}}{\prod_{n_r=1}^{N_r}\prod_{n_t=1}^{N_t}\left(\frac{\bar{\gamma}_{n_t,n_r}}{m}\sum_{\kappa=1}^{K_C}|s_v^{(\kappa)}-s_{v'}^{(\kappa)}|^2+4\right)^m}, \tag{6.14}$$

The derivation of (6.14) is presented in Appendix 6.6.1.

**Proposition 1.** *In the case of 1D constellations, the union bound in (6.14) is a convex function on the convex set*

$$\left\{-(1-e_{vv'})\mathbb{L}+e_{vv'}x' < s_v-s_{v'} < -(1-e_{vv'})x'+e_{vv'}\mathbb{L},\ \forall e_{vv'}\in\{0,1\},\right.$$
$$\left.\forall v,v'\in\{1,\ldots,2^B\}\right\}, \tag{6.15}$$

*where* $\mathbb{L}$ *is a large positive number,* $e_{vv'}$ *is a binary variable and* $x'$ *is given by*

$$x' = \frac{2}{\sqrt{\bar{\gamma}(2+\frac{1}{mN_tN_r})}}, \tag{6.16}$$

*where* $\bar{\gamma}$ *is the total average SNR for each received matrix* $\boldsymbol{G}$.

*Proof.* In Appendix 6.7.1. ∎

In (6.16), when $m\to\infty$ or $N_tN_r$>>1, $x$ tends to $\sqrt{2}/\bar{\gamma}$ and in the limit of high SNR, $x$ tends to zero. The union bound in (6.14) can be upper bounded as

$$P_s \le \frac{1}{2^B}\sum_{v=1}^{2^B}\sum_{\substack{v'=1\\v'\neq v}}^{2^B}\frac{(4m)^{mN_tN_r}}{\mu\left(\sum_{\kappa=1}^{K_C}|s_v^{(\kappa)}-s_{v'}^{(\kappa)}|^2\right)^{mN_tN_r}}, \tag{6.17}$$

where $\mu$ is defined as

$$\mu \coloneqq \prod_{n_r=1}^{N_r} \prod_{n_t=1}^{N_t} \bar{\gamma}_{n_t,n_r}. \tag{6.18}$$

The derivation of (6.17) is presented in Appendix 6.6.2.

**Proposition 2.** *In the case of 1D constellations, the union bound in (6.17) is a convex function on the convex set*

$$\left\{0 < s_{v'} - s_v < \mathbb{L}, \forall v \in \{1, ..., 2^B\}, v' \in \{v+1, ..., 2^B\}\right\}. \tag{6.19}$$

*Proof.* In Appendix 6.7.2. ∎

For the case of the Rayleigh channel, which corresponds to $m = 1$ in the Nakagami-$m$ model, (6.17) can be simplified to

$$P_s \leq \frac{1}{2^B} \sum_{v=1}^{2^B} \sum_{\substack{v'=1 \\ v' \neq v}}^{2^B} \frac{4^{n_t,n_r}}{\mu \left( \sum_{\kappa=1}^{K_C} |s_v^{(\kappa)} - s_{v'}^{(\kappa)}|^2 \right)^{N_t,N_r}}. \tag{6.20}$$

For the AWGN channel, which corresponds to the limiting case in Nakagami-$m$ model with $m \to \infty$, the bound is given by

$$P_s \leq \frac{1}{2^B} \sum_{v=1}^{2^B} \sum_{\substack{v'=1 \\ v' \neq v}}^{2^B} \exp\left( -\frac{N_t N_r}{4N_0} \sum_{\kappa=1}^{K_C} |s_v^{(\kappa)} - s_{v'}^{(\kappa)}|^2 \right). \tag{6.21}$$

A simple proof for (6.21) is presented in Appendix 6.6.3.

If, in each space-time block of the scheme, only one symbol from the multidimensional constellation is transmitted, the SER and BLER of the STBC block become identical. Therefore, the above bound can be used for finding the locally optimum constellations for minimizing the BLER of a space-time block. For the generalized scheme, the SER bound is used to optimize different-sized independent multidimensional constellations used with the scheme, even though this is no longer an appropriate bound on the BLER.

By considering the Hamming distance $D_H(v,v')$ between each pair of constellation points, the corresponding bound on the BER can be written as

$$P_b \leq \frac{1}{B2^B} \sum_{v=1}^{2^B} \sum_{\substack{v'=1 \\ v' \neq v}}^{2^B} \frac{D_H(v,v')(4m)^{mN_tN_r}}{\mu \left( \sum_{\kappa=1}^{K_C} |s_v^{(\kappa)} - s_{v'}^{(\kappa)}|^2 \right)^{mN_tN_r}}. \tag{6.22}$$

For 1D constellations, labels of the constellation points are found by considering the weights $D_H(s,\hat{s})$. Hence, both $s_v > s_{v'}$ and $s_v < s_{v'}$ may happen. If we consider all possibilities for the sign of pairwise differences, $2^{(2^B-1)!}$ different subproblems should be solved. Therefore, the optimization procedure is not possible in polynomial time. In addition, solving all subproblems limits the convexity to the case of $mN_tN_r \in \mathcal{N}$. However, for a specific labelling (e.g., Gray mapping), only one subproblem should be solved. Here, we show that under a specific bit-to-symbol mapping, (6.22) is convex.

**Proposition 3.** *In the case of 1D constellations, for a given bit-to-symbol mapping, the union bound in (6.22) is a convex function on the convex set defined in (6.19).*

*Proof.* It is shown in Appendix 6.7.2 that for 1D constellations (6.17) is a convex function on the convex set defined in (6.19). In the proof, without loss of generality, we assumed $s_1 \leq s_2 \leq ... \leq s_{2^B}$. To keep this condition, the weights $\hat{D}_H(v,v') = D_H(a_v, a_{v'})$ should be used instead of $D_H(v,v')$ in (6.22), where $\mathbf{z}$ is the vector of indices of a given mapping. $\hat{D}_H(v,v')$ is always a non-negative integer and the non-negative weighted sum of convex functions is also a convex function [142]. Hence, (6.22) is convex on (6.19) for a given bit-to-symbol mapping. ■

One of the important factors in deriving bounds for the optimization of constellations is considering their complexity. For example, well-known union bounds on the performance of constellations in the AWGN channel based on Q(·) function in [143] are quite difficult to optimize for medium-to-large constellations since each evaluation of Q(·) takes a relatively long time in comparison to the simplified bounds presented

in (6.17), (6.20), (6.21), and (6.22). Furthermore, in many cases, optimization based on more complex bounds results in very little improvement. Therefore, throughout this chapter, we only optimize (6.17) and (6.22) to find SER-minimizing and BER-minimizing 1D, 2D and multidimensional constellations.

## 6.3 Optimization Criteria and Algorithms

In this section, the optimization problems, optimization procedure and the choice of the method, are discussed. The labelling search algorithm as the second method of finding BER improving constellations is explained. Furthermore, samples of optimized constellations are shown.

### 6.3.1 Optimization Problems

For finding optimized constellations, the union bound given in (6.17) on the BLER is minimized. To improve the performance by employing the shaping instead of increasing the power, the only constraint used in the optimization of multidimensional constellations is that the average power of the constellation points is limited to one. The optimization problem for minimizing BLER is to find $s_v^{(\kappa)}$ for all $\kappa \in \{1, ..., K_C\}$ and $v \in \{1, ..., 2^B\}$ that will

$$\begin{aligned} \text{minimize} \quad & \sum_{v=1}^{2^B} \sum_{v'=v+1}^{2^B} \frac{C_B}{\left(\sum_{\kappa=1}^{K_C} |s_v^{(\kappa)} - s_{v'}^{(\kappa)}|^2\right)^{mN_tN_r}}, \\ \text{subject to} \quad & \frac{1}{2^B K_C} \sum_{v=1}^{2^B} \sum_{\kappa=1}^{K_C} |s_v^{(\kappa)}|^2 \leq 1, \end{aligned} \tag{6.23}$$

where $C_B = 2(4m)^{mN_tN_r}/(2^B\mu)$, which is a constant and does not affect the optimization. In (6.23), $v'$ is started from $v+1$, since $|s_v^{(\kappa)} - s_{v'}^{(\kappa)}|$ and $|s_{v'}^{(\kappa)} - s_v^{(\kappa)}|$ result in an equal PEP. Note that (6.23) does not depend on $\mu$ and therefore the output

of the optimization is an SNR-independent constellation. For the case of the BER optimization, the problem is written as

$$\begin{aligned} \text{minimize} \quad & \sum_{v=1}^{2^B} \sum_{v'=v+1}^{2^B} \frac{D_H(v,v')C'_B}{\left(\sum_{\kappa=1}^{K_C} |s_v^{(\kappa)} - s_{v'}^{(\kappa)}|^2\right)^{mN_tN_r}}, \\ \text{subject to} \quad & \frac{1}{2^B K_C} \sum_{v=1}^{2^B} \sum_{k=1}^{K_C} |s_v^{(k)}|^2 \le 1, \end{aligned} \tag{6.24}$$

where $C'_B = 2(4m)^{mN_tN_r}/(B2^B\mu)$. Due to convexity of the problem for 1D constellations, the following convex programs are used to optimize 1D constellations by minimizing the SER and the BER, respectively:

$$\begin{aligned} \text{Minimize} \quad & \sum_{v=1}^{2^B} \sum_{v'=v+1}^{2^B} \frac{C_B}{(s_v - s_{v'})^{2mN_tN_r}}, \\ \text{subject to} \quad & \frac{1}{2^B} \sum_{v=1}^{2^B} |s_v|^2 \le 1, \\ & s_{v'} - s_v \le \mathbb{L}, s_{v'} - s_v \ge 0, \\ & \forall v \in \{1, ..., 2^B\}, v' \in \{v+1, ..., 2^B\}. \end{aligned} \tag{6.25}$$

$$\begin{aligned} \text{Minimize} \quad & \sum_{v=1}^{2^B} \sum_{v'=v+1}^{2^B} \frac{\hat{D}_H(v,v')C'_B}{(s_v - s_{v'})^{2mN_tN_r}}, \\ \text{subject to} \quad & \frac{1}{2^B} \sum_{v=1}^{2^B} |s_v|^2 \le 1, \\ & s_{v'} - s_v \le \mathbb{L}, s_{v'} - s_v \ge 0, \\ & \forall v \in \{1, ..., 2^B\}, v' \in \{v+1, ..., 2^B\}. \end{aligned} \tag{6.26}$$

In (6.25) and (6.26), due to the normalization of the total constellation power to one, the distance of points cannot be greater than two. As such, we set $\mathbb{L} = 2$. Note that the equalities in constraints are not activated. Otherwise, the objective function tends to infinity.

### 6.3.2 Optimization Procedure and the Choice of the Method

The constellations optimized in this chapter do not need to be updated. Thus, only offline optimization is discussed in this section. For the nonlinear programs (NLPs) in (6.23) and (6.24), two optimization methods, including the interior-point method (IPM) and the sequential quadratic programming method (SQPM), were employed. These two classes of methods are typically used for solving constrained optimization problems. IPM works based on the iterative moving in the interior of the feasible set, determined using the constraints, and decreasing a multiplier until a perturbed Karush-Kuhn-Tucker (KKT) conditions tends to the original KKT. IPM initially walks far from the boundary of the feasible set and iteratively gets closer to the boundary. In each iteration of the SQPM, a quadratic program, which is generated by the quadratic approximation of the objective function, is solved. At each step, the Jacobian and the Hessian are approximated and a step length is determined using a line search in the direction of the minima. Each iteration in IPM typically is more complex than SQPM but fewer iterations are needed to achieve a good solution. For a detailed description of IPM see [144,145], and for SQPM see [129,144] and references therein. In the both cases, the step size decrease as the solver goes closer to a local optima. Thus, the minimum step size is used as the stopping criteria. As the size of the constellation, and consequently the number of variables, increases, we need to increase the maximum number of iterations. As $m$, $N_t$, or $N_r$ increases, the complexity of computation of the Jacobian and Hessian and estimation of the quadratic program increases and the optimization slows down. For all methods, the vector of complex symbols is transferred to the double-size vector of real variables.

For both small and large constellations, e.g., 16-2D and 256-4D, both IPM and SQPM converge to a locally optimum solution with a very similar objective function

value. However, in all cases SQPM converges to the best final value faster. In contrast, IPM finds a solution within ±1% of the final solution faster. As an example, for optimization of a 256-4D SER-minimizing constellation, IPM finds a good solution in the feasibility region in around 51000 steps. This takes around 200 seconds using Matlab on a computer with 24 GB RAM and a 3.40 GHz i7-3770 CPU. In comparison, SQPM finds the same solution in around 160000 steps, which takes around 300 seconds, by setting $m = 1$, $N_t = 1$ and $N_r = 1$. Although further optimization is not effective on the performance of the system, SQPM and IPM converge to the best achievable solution in around 410000 and 670000 steps, corresponding to around 700 and 2200 seconds, respectively. In fact, since IPM can move far from the boundary of the feasible set, it converges faster to a good solution. In the long run, since the computation of each step in SQPM is cheaper, SQPM can move deeper in the feasible set during a limited time and generates slightly better results. Nevertheless, for large constellations, IPM finds a good solution faster than SQPM. Therefore, it is used as the preferred optimization method in this chapter.

To improve the results, we restart the solver from a slightly perturbed starting point sequentially to find several locally minimum solutions and we choose the best of these. For generating new starting points, small perturbation coefficients from $\mathcal{CN}(0, 0.2)$ are randomly generated and added to the initial starting point. Then the new starting point total power is normalized to satisfy the power constraint in (6.23) and (6.24).

In addition to IPM and SQPM, we also examined simulated annealing in [146] and genetic algorithm as two well known methods of global optimization. For the case of an simulated annealing algorithm, although improvement in bound value is observable, it converges very slowly and the results are worse than the solution found using IPM and SQPM. For genetic algorithm, the results were even worse as it does not find any useful solution. Indeed, due to the continues nature of the feasible set,

evolutionary algorithms cannot find good solutions.

The convex problems in (6.25) and (6.26) can be modelled and solved using cvx. Due to energy efficiency, optimum constellations for coherent systems have zero mean. It has been shown that they are typically symmetric around zero [53]. Therefore, as a good solution, we can optimize the constellation for the positive points alone, i.e., $s_{2^{(B-1)}}, ..., s_{2^B}$. Therefore, we only have $2^{(B-1)}$ free variables to optimize. Thus, for optimization of a 1D constellation with $B = 6$, only 32 variables should be optimized. For the case of the BER-minimizing constellation, we set the bit-to-symbol mapping to Gray. Gray mapping is optimal at high SNR for regular constellations in the AWGN channel [53]. Here, we assume it remains good in the Nakagami-$m$ channel. For the sake of comparison, we also used IPM to solve the equivalent 1D problems in (6.23) and (6.24) to find locally optimal solutions. The result shows that the convex optimization of a 16-1D constellation, can provide up to 0.2 dB better results than non-convex optimization.

To initiate the solver for optimizing based on the SER or BLER with a good starting point, all constellations are initially selected from cubic constellations. For example, the rectangular QAM constellations are used as the initial point for optimization of 2D constellations. To initiate the optimization for minimizing the BER, the Cartesian product of Gray-mapped PAM constellations is employed.

For optimization in the AWGN channel, a value of $m = 10$ is used in optimization problems (6.23) and (6.24) instead of a very large $m$, since large values of $m$ slows down the optimization procedure and Nakagami-$m$ fading with $m = 10$ is close enough to the AWGN channel. Alternatively, the bound (6.21) can be used for the optimization of constellations in the AWGN channel. However, the result of optimization with this bound does not show good performance since the SNR knowledge is necessary for finding good constellations.

### 6.3.3 Two-Step Optimization of BER-minimizing Constellations

Traditionally, to optimize the constellation for minimizing the BER, a two-step process is used. First an optimum constellation is found based on the shaping metric; and second, the bit-to-symbol mapping is optimized by using an appropriate metric [129]. Therefore, two independent steps are needed to find an optimum constellation for minimizing the BER. For example, in our case, the bound (6.17) on the SER is used to find a constellation with a good shape and then by using an appropriate method, such as the binary switching algorithm in Algorithm 1, the best bit-to-symbol mapping is found. To find the bit-to-symbol mapping, we modify the binary switching algorithm to adapt it to our problem. In Algorithm 1, the cost of each symbol, $c_p(v)$, can be calculated as

$$c_p(v) = \sum_{\substack{v'=1 \\ v' \neq v}}^{2^B} \frac{D_H(v,v')C'_B}{\left(\sum_{\kappa=1}^{K_C} |s_v^{(\kappa)} - s_{v'}^{(\kappa)}|^2\right)^{mN_tN_r}}. \tag{6.27}$$

In Section 6.4, we compare the result of the two-steps optimization method with the constellations achieved using the bound (6.22) on the BER which corresponds to the joint optimization of constellation shaping and bit-to-symbol mapping. We show that the constellations achieved using the optimization of (6.22) outperform the constellations achieved using the two-step method.

### 6.3.4 Samples of Optimized Constellations

By optimizing problems of Section 6.3.1, constellations with improved performance in comparison to the PAM and QAM constellations are achieved. Fig. 6.1 shows samples of 16-1D constellations optimized by solving the problem (6.23) for the Nakagami-$m$ fading channel. We observe that while for the AWGN channel, approximately

equidistant PAM constellations are known to outperform other constellations, the optimal shape is quite different for other cases including Nakagami-$m$ with $m = 3$ and Rayleigh fading.

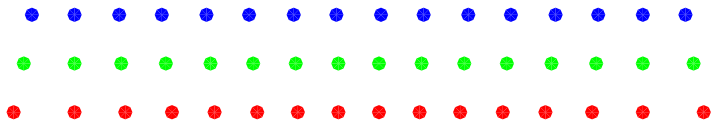


**Figure 6.1:** Comparison of 1D 16-PAM SER optimized constellations for the AWGN channel (top, in blue), the Nakagami-$m$ channel with $m = 3$ (middle, in green), and the Rayleigh channel, i.e., Nakagami-$m$ with $m = 1$ (bottom, in red).

The best known 2D constellations at high SNR for minimizing the SER are Voronoi constellations, where signal points are positioned approximately on a hexagonal grid which we refer to as the HEX constellations [147–149] or penny packing [150]. Fig. 6.2(a) shows a sample of 16-2D constellations optimized by solving the problem (6.23). Interestingly, the optimized constellation for the AWGN channel is HEX-like while the one for the Rayleigh channel is on two polygons (one inside of the other) with a zero amplitude point in the middle. Fig. 6.2(b) shows the samples of the optimized 2D constellations for minimizing the BER by solving the problem (6.24). Interestingly, for the AWGN channel, the optimized constellation is a HEX-like one, while for the Rayleigh channel it is only slightly different from a 16-QAM constellation.

QAM constellations with order $2^B, B \in 3, 5, 7, ..., 2n + 1$, are not energy efficient. However, by solving (6.23), energy efficient alternatives can be generated. Fig. 6.2(c) and Fig. 6.2(d) illustrates the 8-2D constellations optimized for the SER and BER of the AWGN and Rayleigh fading channels, respectively.

The two 2D projections of a sample optimized 16-4D constellation is plotted in Fig. 6.3. In this figure, each 2D constellation point represents two dimensions of a

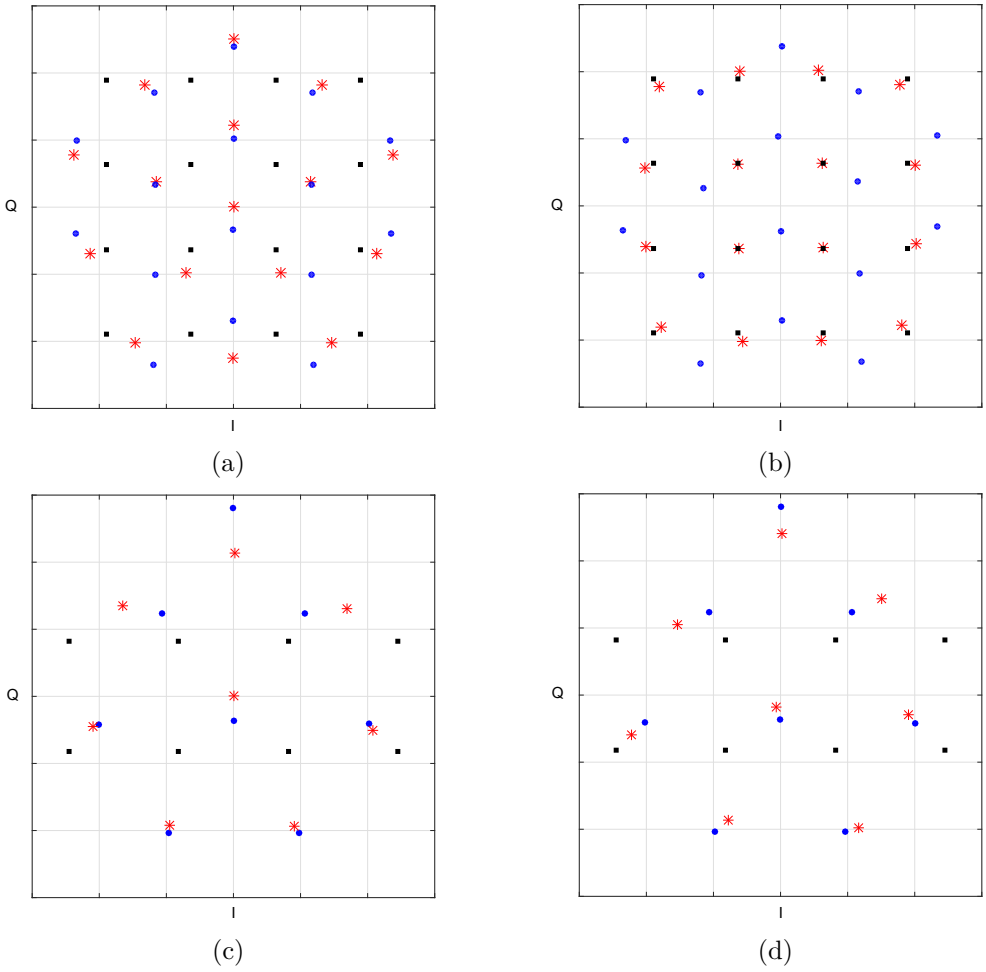



**Figure 6.2:** Comparison of a) 16-QAM and 16-2D SER optimized constellations, b) 16-QAM and 16-2D BER optimized constellations, c) 8-QAM and 16-2D SER optimized constellations and d) 8-QAM and 16-2D BER optimized constellations. The QAM constellations are shown with black squares and the optimized constellations for AWGN and the Rayleigh fading channel are shown with blue circles and red asterisks, respectively.

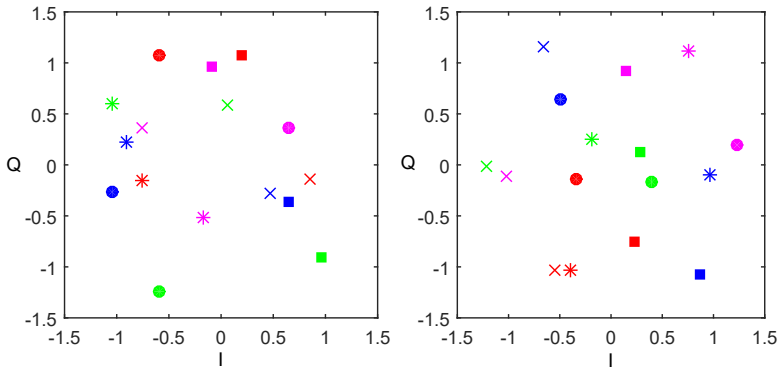



**Figure 6.3:** Sample of the optimized 16-4D constellation used for the scheme $\mathbf{G}_2$.

4D constellation point, and points with the same label are indicated with the same marker and colour. This figure shows that constellation points generated by solving (6.23) can be irregularly placed anywhere in the signal space.

## 6.4 Numerical Results and Discussions

The performance of the scheme in uncoded systems and in presence of polar codes is evaluated in this section.

### 6.4.1 The Uncoded Scheme

In this section, we evaluate the performance of the constellations optimized in Section 6.3 in comparison to the best-known constellations in the literature. The channel is modelled as experiencing uncorrelated Nakagami-$m$ fading with AWGN. The results of SISO and MIMO antenna configurations are discussed in the first and the second parts of this section, respectively.

#### 6.4.1.1 Results for the SISO Configuration

In this section, the performance of constellations optimized for the SISO configuration are evaluated. For comparison, constellations were also optimized by using a bound on the SER of the SISO AWGN channel in [148, 151, 152]. The corresponding optimization problem can be written as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2^B\sqrt{2\pi}} \sum_{v=1}^{2^B} \sum_{\substack{v'=1 \\ v' \neq v}}^{2^B} \frac{\exp\left(-\frac{1}{4N_0}|s_v - s_{v'}|^2\right)}{\frac{1}{\sqrt{2N_0}}|s_v - s_{v'}|^2}, \\ \text{subject to} \quad & \frac{1}{2^B} \sum_{v=1}^{2^B} |s_v|^2 \leq 1. \end{aligned} \tag{6.28}$$

For the optimization of 2D constellations based on (6.28), the SNR is set to 20 dB (the SNR value is required for this optimization and we found the best results at the mentioned value).

In Fig. 6.4, the BLER of 64-1D constellations has been evaluated in the Nakagami-$m$ channel with different $m$ values and for the spectral efficiency of 6 bpcu. In each case, the performance of the optimized constellation for the corresponding $m$-factor is compared with the approximately equidistant 64-PAM constellation which is the well known capacity maximizing high-SNR 1D constellation for the AWGN channel. In the Nakagami-$m$ channel with $m = 1$, which is the Rayleigh channel, the constellation optimized for $m = 1$ shows a 0.4 dB gain in comparison to 64-PAM at a BLER of $10^{-4}$. For the case of $m = 3$, the constellation optimized for $m = 3$ shows a 0.2 dB gain in comparison to equidistant 64-PAM at a BLER of $10^{-4}$, and for the case $m \to \infty$, which is the AWGN channel, the constellation optimized for $m = 10$ performs approximately the same as 64-PAM.

Fig. 6.5 shows the comparison of the BLER of 2D and 4D constellations in the AWGN channel for 4 bpcu. The length of the block is considered to be two channel uses since the size of the largest constellation is 4D in these figures. These results

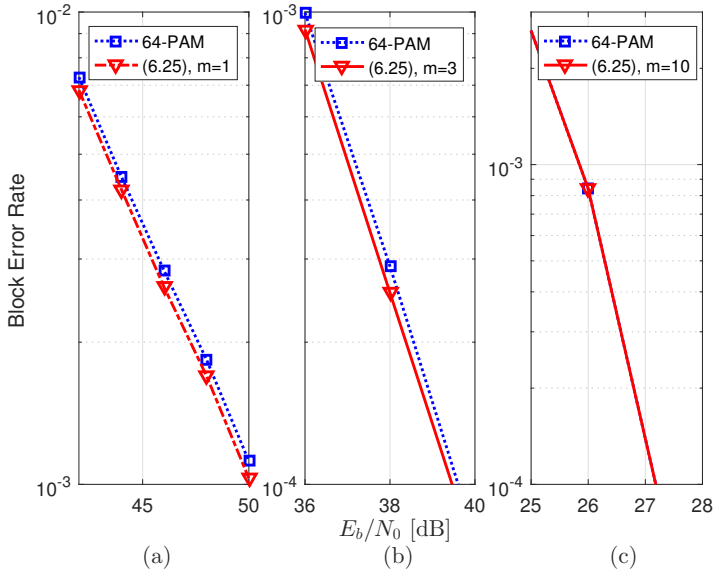



**Figure 6.4:** BLER comparison of a 1D constellation with a SISO configuration for 6 bpcu in Nakagami-$m$ channels with different values of $m$, a) equidistant 64-PAM vs. 64-PAM optimized for $m = 1$ in Nakagami-$m$ fading with $m = 1$, b) equidistant 64-PAM vs. 64-PAM optimized for $m = 3$ in Nakagami-$m$ fading with $m = 3$ , c) equidistant 64-PAM vs. 64-PAM optimized for $m = 10$ in an AWGN channel.

indicate that the 16-2D constellations, optimized either by solving problem (6.23) by setting $m = 10$ or by solving problem (6.28), perform similarly to each other and have approximately the same performance as the 16-2D constellation optimized in [127] or a 16-HEX constellation. Furthermore, all these 16-2D optimized constellations work 0.4 to 0.5 dB better than 16-QAM. The 256-4D constellation optimized in [127] performs around 0.2 dB better than optimized 16-2D constellations, while the 256-4D constellation optimized by solving problem (6.23) with $m = 10$ shows an additional 0.2 dB gain. Moreover, the performance of some of these constellations were checked in the Rayleigh channel and, as expected, the 16-2D constellation designed for Rayleigh outperforms the 2D constellation optimized for AWGN by 0.2 dB, and the 16-QAM constellation by 0.4 dB at a BLER of $10^{-5}$.

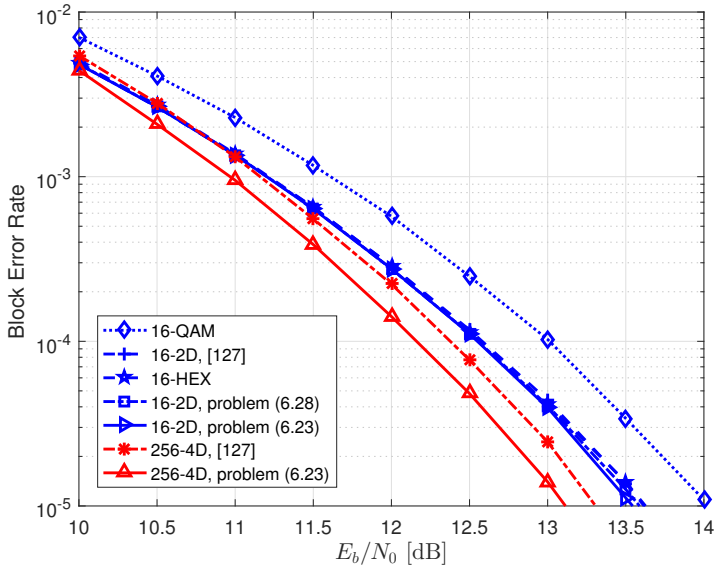



**Figure 6.5:** BLER comparison of a 2D constellation with a SISO configuration for 4 bpcu in an AWGN channel.

Constellations optimized for minimizing the BER can be achieved by solving problem (6.23) on the SER and finding the best bit-to-symbol mapping by using the binary switching algorithm as described in Section 6.3, which we refer to as method A, or alternatively by solving problem (6.24) on the BER which is the joint optimization of the constellation shape and the bit-to-symbol mapping, which we refer to as method B. Here, we compare these two methods. Fig. 6.6 illustrates the BER comparison of different 16-2D constellations in the SISO Rayleigh channel. It shows that the 16-2D constellation constructed based on method A by solving the problem (6.23) with $m = 1$ outperforms the 16-2D constellation constructed based on method A by using the constellation optimized in [127] for 0.4 dB at a BLER of $10^{-5}$; the 16-2D constellation optimized based on method B with $m = 1$ provides an additional 0.4 dB gain.

For the case of the AWGN channel where we used a constellation optimized for

the AWGN channel, there is only a small preference in performance for the constellation optimized using method B in comparison to method A, since as we observe in Fig. 6.2(a) and Fig. 6.2(b), constellations optimized for SER using bound (6.17) and for BER using bound (6.22) already have quite similar HEX-like shapes.

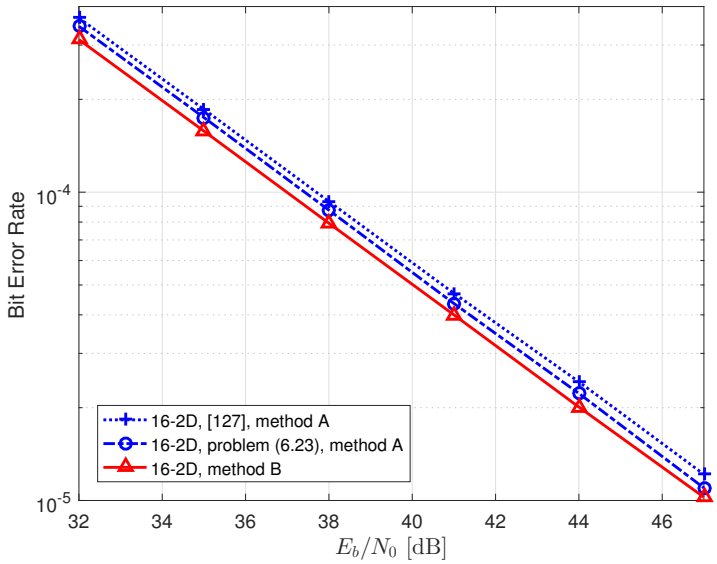



**Figure 6.6:** BER comparison of 2D constellations with a SISO configuration for 4 bpcu in a Rayleigh channel.

In Fig. 6.7, the BER performance of the 2D constellation optimized by solving problem (6.24) under a SISO antenna configuration and in the AWGN channel is examined. Among 2D constellations, 16-QAM with Gray mapping is outperformed by the 16-2D constellation optimized for Nakagami-$m$ with $m = 10$ by 0.3 dB at a BLER of $10^{-5}$. Furthermore, the 256-4D constellation optimized for Nakagami-$m$ with $m = 10$ outperforms the optimized 16-2D constellation by 0.5 dB at a BLER of $10^{-5}$.

The above results show that by increasing the dimensionality of the constellation, there is more space for points to be further apart and this improves performance. Furthermore, by increasing the number of points of a constellation the performance gap

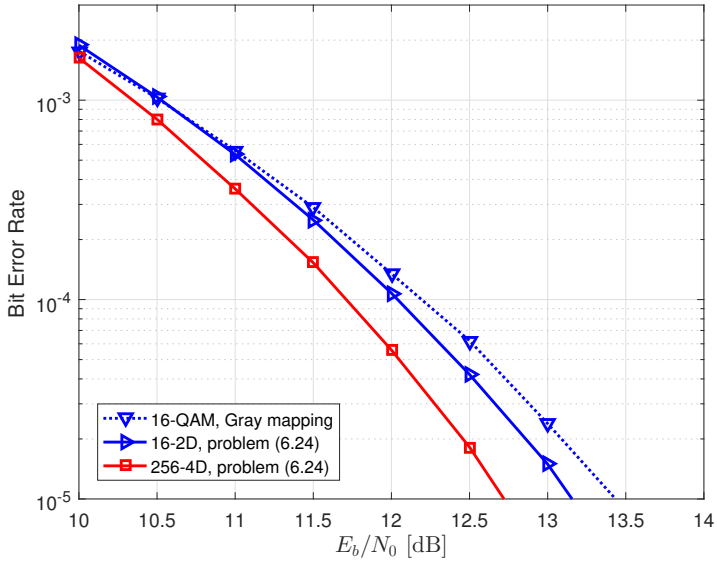



**Figure 6.7:** BER comparison of 2D constellation with a SISO configuration for 4 bpcu in an AWGN channel.

between constellations optimized based on the bounds (6.17) and (6.22) and distance based constellations (e.g. [127]) increases. Results of this section are summarized in Table 6.1.

#### 6.4.1.2 Results for the MIMO Configuration

In this part of the performance evaluation, as the baseline, OSTBCs with QAM constellations are compared against the schemes $\mathbf{G}_2$-$\mathbf{G}_4$, which consists of the use of multidimensional constellations with OSTBC. Since BLER (or the BER of a block) is used in the comparisons, constellations were optimized by using the problems (6.23) and (6.24), and; therefore, one multidimensional constellation is used in each space-time code block. As a reference for comparison, the same scheme is constructed by using OSTBC and the constellation proposed in [127]. The scheme is also compared with the Golden code [35] and the algebraic multiple-input single-output (MISO)

**Table 6.1:** Performance advantage of optimized constellations in comparison with the best-known constellations in the literature for the SISO system.

| Constellation | $m$ | bpcu | Metric | Performance Advantage |
|---|---|---|---|---|
| 64-1D | 1 | 6 | BLER | 0.4 dB |
| 64-1D | 3 | 6 | BLER | 0.2 dB |
| 64-1D | 10 | 6 | BLER | 0 dB |
| 16-2D | 10 | 4 | BLER | $\sim$ 0 dB |
| 256-4D | 10 | 4 | BLER | 0.2 dB |
| 16-2D | 10 | 4 | BER | 0.4 dB |
| 256-4D | 10 | 4 | BER | 0.8 dB |

code in [139] which we refer to as the "Oggier code". The constellations used with the Golden code for 2 bpcu and 4 bpcu are BPSK and QPSK, respectively, and with the Oggier code for 1 bpcu is BPSK. Furthermore, the scheme is compared with the QOSTBC with optimal rotation in [56], used with QPSK for a spectral efficiency of 2 bpcu.

In Fig. 6.8 the BLER performance of the Golden code, Alamouti's OSTBC $\mathbf{G}_1$ and the scheme $\mathbf{G}_2$, all with 2 bpcu, are compared in a Rayleigh channel. This result shows that scheme $\mathbf{G}_2$ with a constellation optimized by solving the problem (6.23) has the same performance as scheme $\mathbf{G}_2$ with a constellation optimized in [127]. Both schemes outperform OSTBC by 0.4 dB in a $2 \times 1$ configuration and by 0.5 dB in a $2 \times 2$ configuration at a BLER of $10^{-4}$. Furthermore, the Golden code in a $2 \times 2$ configuration shows a BLER worse than OSTBC. Indeed, most algebraic codes are designed for high rates and therefore show poor performance at low rates since not all their degrees of freedom are well exploited.

Fig. 6.9 shows the performance comparison of scheme $\mathbf{G}_2$ with OSTBC in $2\times1$ and

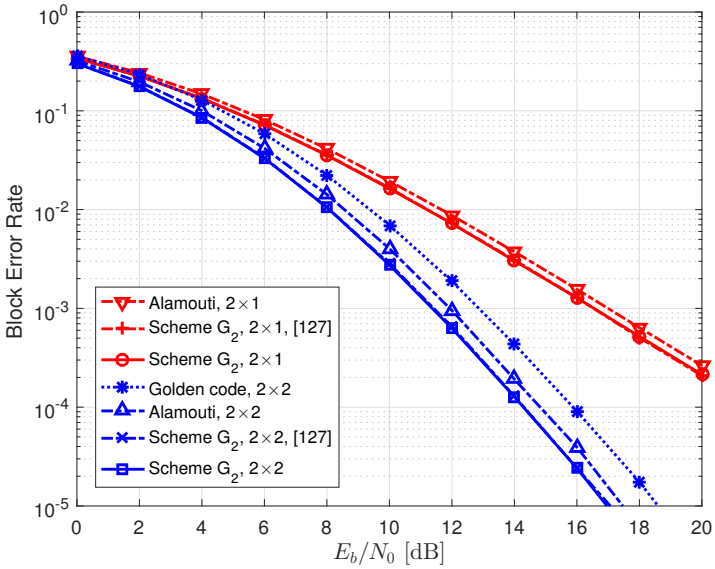



**Figure 6.8:** BLER comparison of scheme $\mathbf{G}_2$, OSTBC and the Golden code for 2 bpcu.

$2 \times 2$ configurations and the Golden code $2 \times 2$ for 4 bpcu in a Rayleigh channel. The outcome indicates that the scheme $\mathbf{G}_2$ with a constellation optimized by solving the problem (6.23) and with a constellation optimized in [127] perform the same. They also work better than OSTBC as one of the best codes for the $2 \times 1$ configuration, by 0.9 dB at a BLER of $10^{-3}$. Furthermore, for the $2 \times 2$ antenna configuration, it is 0.9 dB better than OSTBC and only 0.5 dB worse than the Golden code at $10^{-4}$. Note that the Golden code $2 \times 2$ outperforms $\mathbf{G}_2$ since it benefits from more degrees of freedom, but it also has four times more complexity in terms of complex multiplications in each search for ML decoding even though the number of searches is the same as that in the scheme.

Fig. 6.10 shows the error performance of the schemes $\mathbf{G}_3$ and $\mathbf{G}_4$ for $3 \times 1$ and $4 \times 1$ antenna configurations, respectively, in a Rayleigh channel. The schemes $\mathbf{G}_3$ and $\mathbf{G}_4$ have approximately the same performance when used with either the constellation

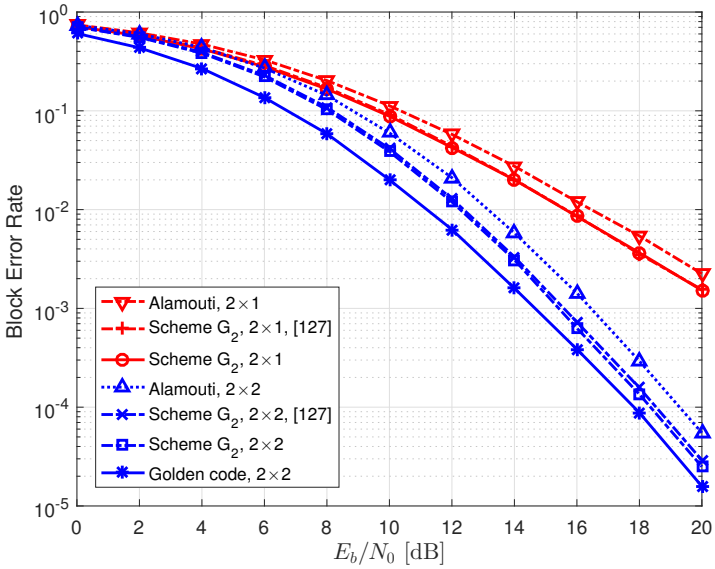



**Figure 6.9:** BLER comparison of the scheme $\mathbf{G}_2$, OSTBC and the Golden code for 4 bpcu.

optimized by solving the problem (6.23) or the constellation optimized in [127], and they outperform OSTBC by 1.5 dB in $3\times1$ and in $4\times1$ configurations at a BLER of $10^{-4}$. Furthermore, the BLER comparison of the scheme and OSTBC in $3\times2$ and $4\times2$ configurations in Fig. 6.11 shows 1.7 dB and 1.8 dB improvement at $10^{-4}$, respectively. To compare the scheme with algebraic codes, the recently designed MISO code in [139] (the "Oggier code") that can support lower rates was tested; similar to the Golden code in Fig. 6.8, its BLER is worse than the corresponding OSTBC.

Fig. 6.12 shows the performance of the scheme $\mathbf{G}_3$ in comparison to OSTBC and QOSTBC $4\times1$ and $4\times2$ for 2 bpcu. The results show that the scheme outperforms OSTBC by around 4 dB at $10^{-3}$ and also outperforms QOSTBC $4\times1$ and $4\times2$ by 0.8 dB and 0.4 dB at $10^{-4}$, respectively. Note that the improvement in comparison to OSTBC or QOSTBC is achieved at the expense of more decoding complexity. In the case of QOSTBC, joint pairwise decoding results in a lower number of searches,

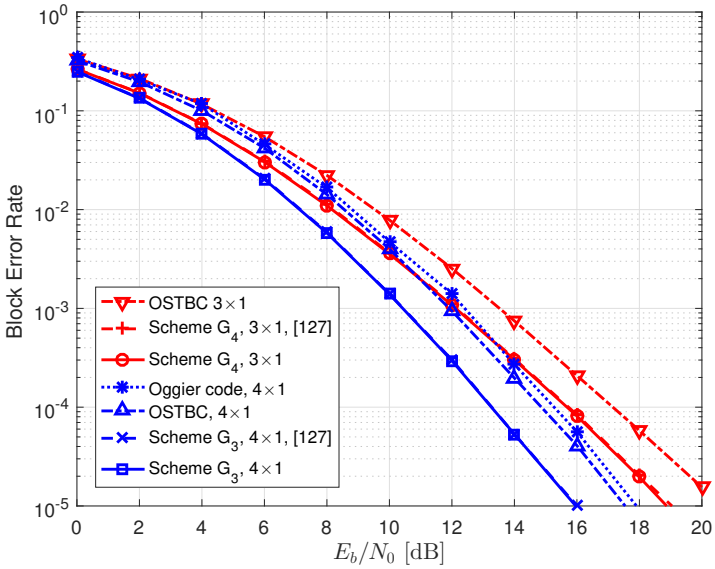



**Figure 6.10:** BLER comparison of the schemes $\mathbf{G}_3$ and $\mathbf{G}_4$, OSTBC and the Oggier code for 1 bpcu, with $N_r = 1$.

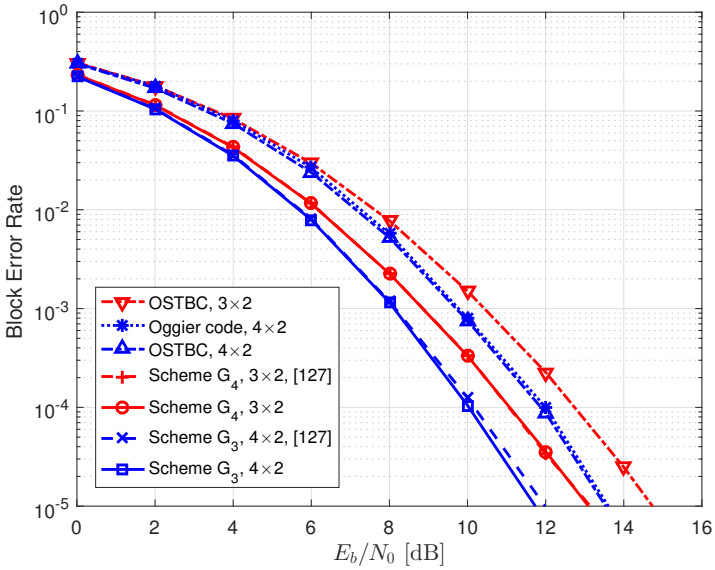



**Figure 6.11:** BLER comparison of the schemes $\mathbf{G}_3$ and $\mathbf{G}_4$, OSTBC and the Oggier code for 1 bpcu, with $N_r = 2$.

but each search is more complex than the decoding of the scheme. Furthermore, unlike the previous figures, the performance of the scheme with the constellation achieved from solving the problem (6.23) outperforms the scheme with a constellation optimized in [127] by 0.3 dB and 0.2 dB for $4\times1$ and $4\times2$ configurations, respectively. In comparison to [127], since the modulation is optimized for the Rayleigh fading channel, the performance is improved.

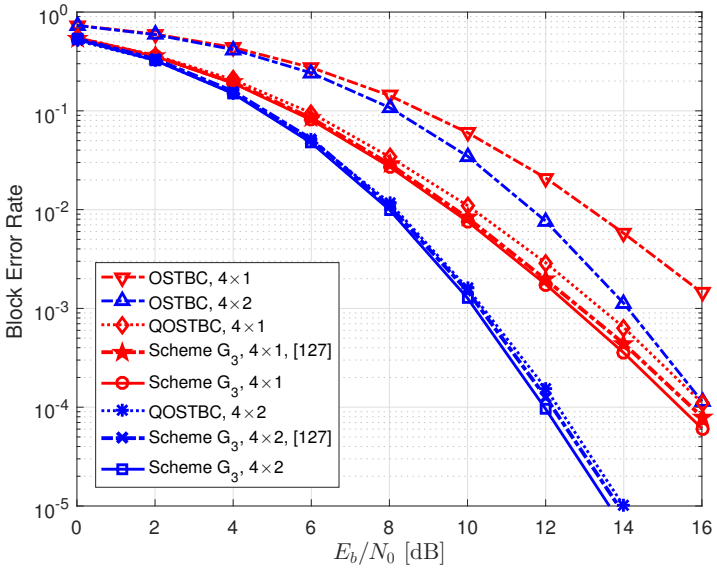



**Figure 6.12:** BLER comparison of the scheme $\mathbf{G}_3$, OSTBC and QOSTBC for 2 bpcu.

The BER of the scheme $\mathbf{G}_3$ in comparison to the QOSTBC for $4\times1$ and $4\times2$ configurations is shown in Fig. 6.13. For a $4\times1$ configuration, scheme $\mathbf{G}_3$ with a constellation optimized by solving the problem (6.23) outperforms scheme $\mathbf{G}_3$ with method A and a constellation optimized in [127]. It also outperforms QOSTBC by 0.3 dB at BER of $10^{-5}$. For a $4\times2$ configuration, the scheme $\mathbf{G}_3$ with a constellation optimized by solving the problem (6.23) and QOSTBC perform approximately the same and outperform scheme $\mathbf{G}_3$ with method A and a constellation optimized in [127] by 0.5 dB at BER of $10^{-5}$.

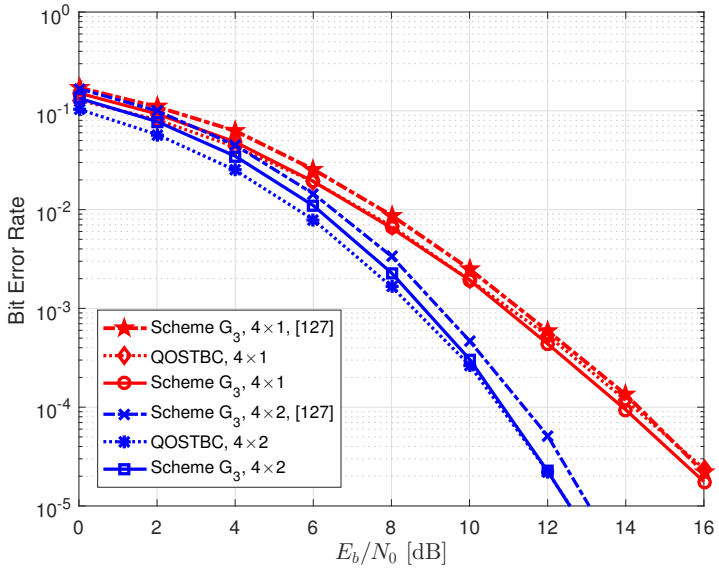



**Figure 6.13:** BER comparison of the scheme $\mathbf{G}_3$ and QOSTBC for 2 bpcu.

Finally, Fig. 6.14 shows the performance of the generalized scheme $\mathbf{G}_5$ with one 16-4D constellation and a QPSK constellation in comparison with the scheme $\mathbf{G}_4$ for a $3 \times 1$ antenna configuration used with a 64-6D constellation and an OSTBC with a QPSK constellation in a Rayleigh channel. We observe that the scheme $\mathbf{G}_4$ outperforms $\mathbf{G}_5$ by 0.4 dB and $\mathbf{G}_5$ outperforms OSTBC by 0.3 dB at the BLER of $10^{-4}$. As explained in Section 6.1, the difference of the generalized scheme $\mathbf{G}_5$ and scheme $\mathbf{G}_4$ can be explained by using a complexity-performance trade-off. For the ML decoding of scheme $\mathbf{G}_4$, 64 searches in 6D space are necessary while for the generalized scheme $\mathbf{G}_5$, only 16 searches in 4D space and 4 searches in 2D space are necessary. Thus, the generalized scheme has a lower decoding complexity and since the points have less space to be far apart, the performance is degraded.

By designing multidimensional constellations adapted to OSTBC, high performance improvements can be achieved at the expense of increasing the complexity

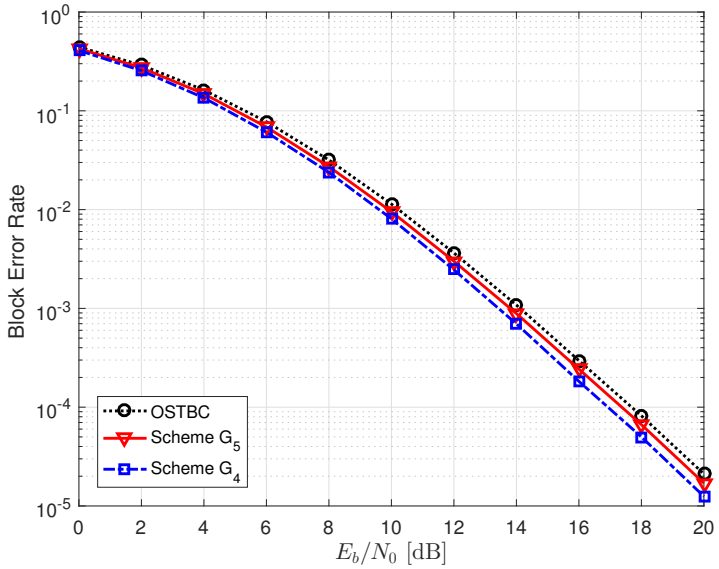



**Figure 6.14:** BLER comparison of the generalized scheme and OSTBC for 1.5 bpcu.

of decoding. Increasing the dimensionality of the constellation improves the performance even though the size of the employed OSTBC, including the number of antennas and the number of time slots, may be increased. Indeed, this is the main reason that schemes $\mathbf{G}_3$ and $\mathbf{G}_4$ provide more gain in comparison to $\mathbf{G}_2$. Additionally, as observed in the results of the generalized scheme shown in Fig. 6.14, there exists a complexity-performance trade-off when part of the independent symbols of the OSTBC are used to carry dependent dimensions of multidimensional constellations. Performance advantages of the results of this section are summarized in Table 6.2 for 2 bpcu. Note that for 1 bpcu, the gain in comparison to the best-known constellations is approximately zero.

**Table 6.2:** Performance advantage of optimized constellations in comparison with the best-known constellations in the literature for the MIMO Rayleigh fading channel and 2 bpcu.

| Scheme | Constellation | Metric | Performance Advantage |
|---|---|---|---|
| $\mathbf{G}_2$, $2\times1$ | 16-4D | BLER | 0 dB |
| $\mathbf{G}_2$, $2\times2$ | 16-4D | BLER | 0 dB |
| $\mathbf{G}_3$, $4\times1$ | 256-6D | BLER | 0.3 dB |
| $\mathbf{G}_3$, $4\times2$ | 256-6D | BLER | 0.2 dB |
| $\mathbf{G}_3$, $4\times1$ | 256-6D | BER | 0.5 dB |
| $\mathbf{G}_3$, $4\times2$ | 256-6D | BER | 0 dB |

The performance of $\mathbf{G}_4$ for a $3\times1$ antenna configuration under imperfect channel estimation in a Rayleigh fading channel is evaluated in Fig. 6.15. The channel estimation method is assumed to be linear minimum mean square error (LMMSE). As described in [153, 154], the variance of the channel estimation error with LMMSE can be modelled with a factor $\sigma_E^2$ ranging from zero to one that corresponds to a coherent receiver (perfect channel estimation) when it is set to zero and a non-coherent receiver (no channel estimation) when it is set to one. It can be seen that when the channel is estimated imperfectly, the BLER curves show an error floor at relatively high values.

When the regular constellations such as QAM constellations are used, the decision regions are very regular. Therefore, designing a low-complexity ML decoder for these constellations is possible using the simple decision thresholds. However, for the irregular constellations, the decision regions are very complex and designing a decoder based on these regions may be infeasible. In both cases, sphere decoders may decrease the decoding complexity [155]. However, for regular constellations, sphere decoders with lower complexity can be designed.

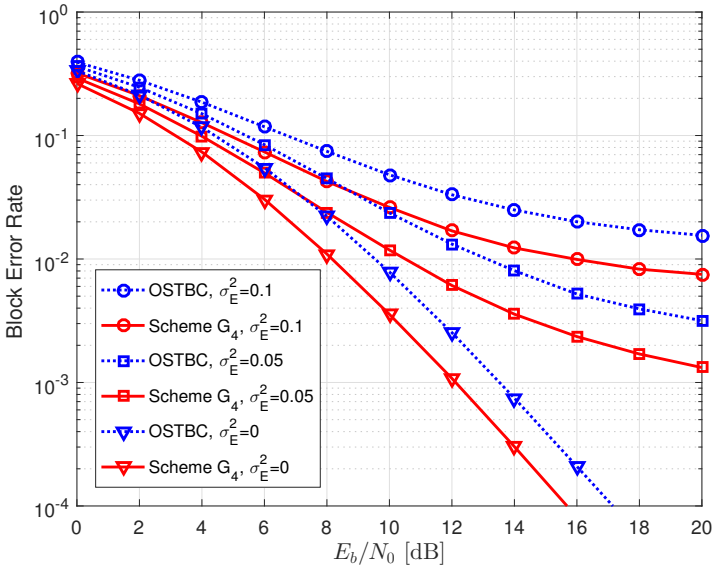



**Figure 6.15:** BLER comparison of the scheme $\mathbf{G}_4$ in three different values of $\sigma_E^2$ for 2 bpcu with a $3 \times 1$ antenna configuration.

### 6.4.2 The Polar Coded Scheme

In this section, the constellation optimized based on the error bounds are used in construction of MLPCM. For all cases, the polar code design method described in Section 4.5 is employed. Fig. 6.16 shows the result of the comparison of MLPCM scheme constructed using 16-HEX, 16-QAM, 8-QAM, circular 8-QAM, and 8-HEX in an AWGN channel with a SISO antenna configuration. All constellations are used with SPM. For MLPCM with 16-point and 8-point constellations $N_{tot}$ is set to 1024 and 768 bits, respectively. The polar codes are constructed at a FER of 0.01. We observe that 16-QAM outperforms 16-HEX and 8-HEX outperforms 8-QAM and circular 8-QAM constellations.

In Fig. 6.17, the performances of the optimized 16-2D constellation for $m = 1$, 16-QAM, circular 8-QAM, optimized 8-2D constellation, and 8-HEX in an independent Rayleigh fading channel with a SISO antenna configuration are evaluated. We observe

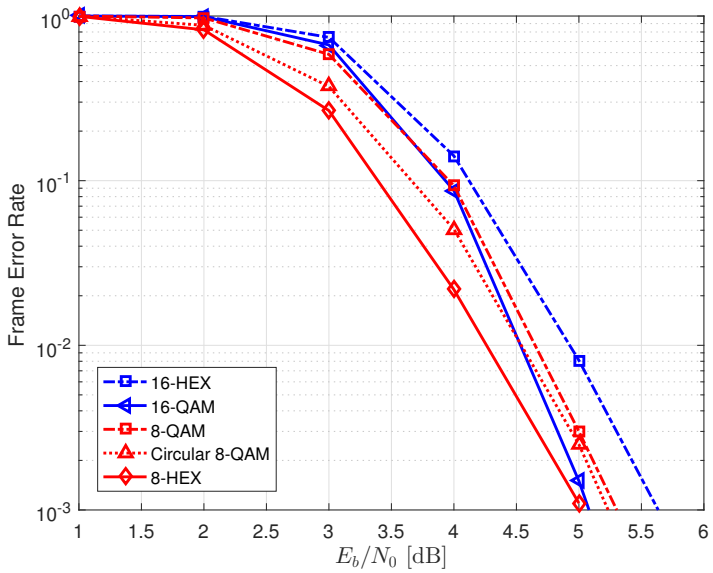



**Figure 6.16:** FER comparison of the MLPCM scheme constructed using 16-HEX, 16-QAM, 8-QAM, circular 8-QAM and 8-HEX in an AWGN channel.

that 8-HEX provides the best performance among 8-point constellations and 16-QAM provides better performance in comparison to the optimized 16-2D constellation for $m = 1$ when used with polar codes at a FER of 0.01.

Fig. 6.18 shows the comparison of the mutual information of 16-HEX, 16-QAM, 8-QAM, circular 8-QAM and 8-HEX constellations in an AWGN channel with a SISO antenna configuration. We observe that the optimized 16-HEX and 8-HEX constellations have higher mutual information than the same size QAM constellations. However, in Fig. 6.16, the performance of the 16-QAM constellation was better than the 16-HEX constellation. Indeed, the higher mutual information does not guarantee better performance when constellations are used with outer FEC codes.

Fig. 6.19 shows the comparison of the mutual information of optimized 16-2D constellation for $m = 1$, 16-QAM, circular 8-QAM, optimized 8-2D constellation for $m = 1$ and 8-HEX in an independent Rayleigh fading channel with a SISO antenna

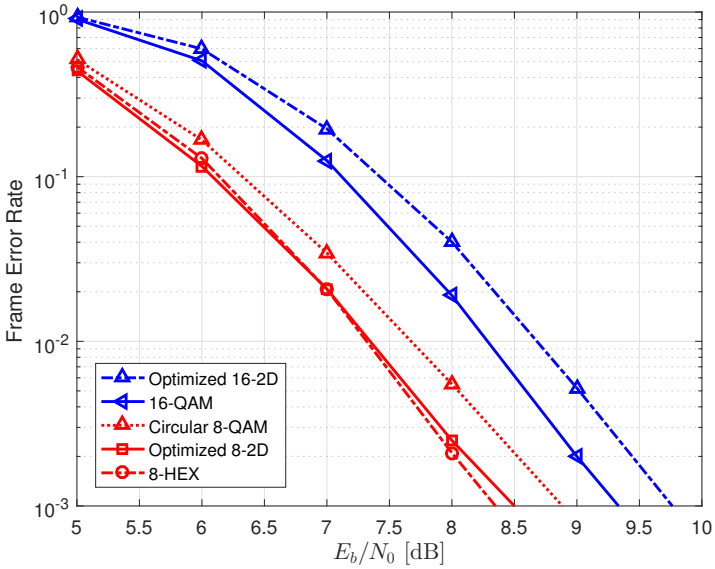



**Figure 6.17:** FER comparison of the MLPCM scheme constructed using optimized 16-2D constellation for $m = 1$, 16-QAM, circular 8-QAM, optimized 8-2D constellation for $m = 1$ and 8-HEX in an independent Rayleigh fading channel.

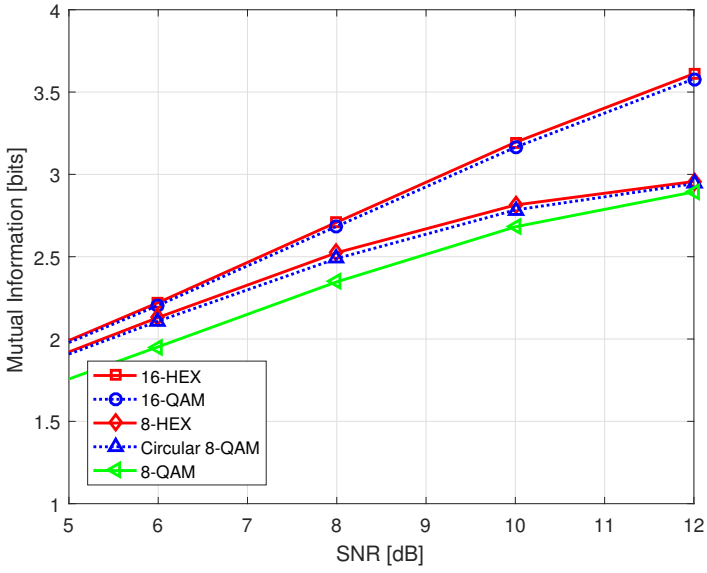



**Figure 6.18:** Comparison of the mutual information of 16-HEX, 16-QAM, 8-QAM, circular 8-QAM and 8-HEX constellations in an AWGN channel.

configuration. We observe that the mutual information of the optimized 16-2D and 8-2D constellations for $m = 1$ is higher than the 16-QAM and 8-HEX constellations, respectively. However, similar to the AWGN channel, 16-QAM and 8-HEX constellations work better when used with polar codes.

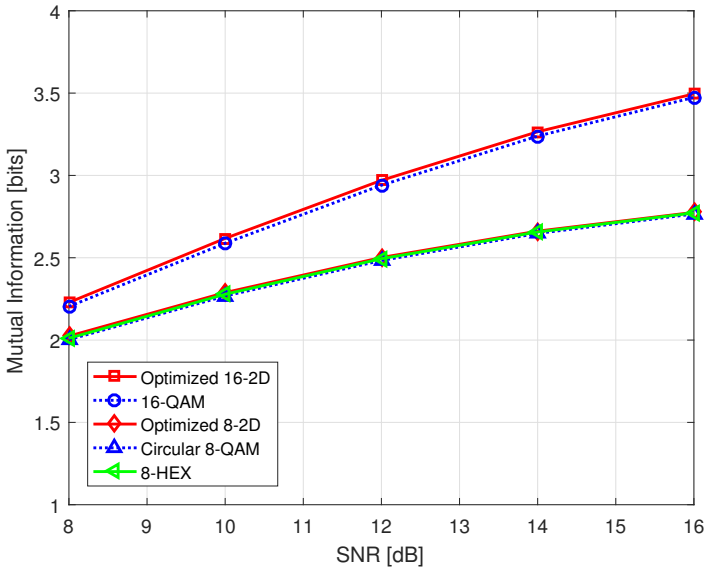



**Figure 6.19:** Comparison of the mutual information of optimized 16-2D constellation for $m = 1$, 16-QAM, circular 8-QAM, optimized 8-2D constellation for $m = 1$ and 8-HEX in an independent Rayleigh fading channel.

Fig. 6.20 provides a comparison of MLPCM constructed using QOSTBC with a BPSK constellation and the OSTBC with a 16-6D constellation at 0.75 bpcu. The channel is a Rayleigh slow fading with $f_d = 10^{-5}$. In this case, since the 16-point cubic constellation cannot be constructed, a 16-2D constellation optimized using the methods of this chapter is employed. For the QOSTBC, the corresponding space-time constellation is constructed by substitution of BPSK in the STBC structure with the corresponding optimal rotation. All codewords are of a total length of 2048 bits. In both $4 \times 1$ and $4 \times 2$ antenna configurations, OSTBC with 16-6D outperforms the QOSTBC for 0.6 dB at a FER of 0.01.

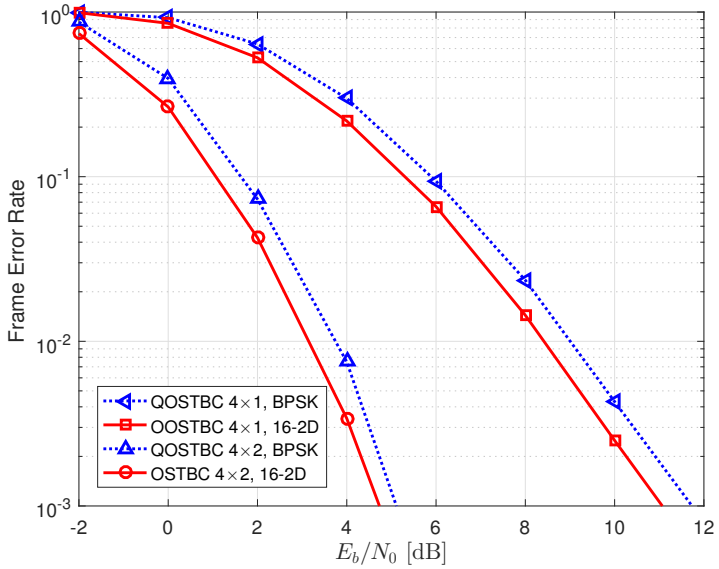



**Figure 6.20:** FER comparison of the MLPCM scheme constructed using QOSTBC with BPSK and OSTBC with a 16-6D constellation in a Rayleigh slow fading channel.

Fig. 6.21 shows the comparison of MLPCM constructed using QOSTBC with QPSK and the optimal rotation and the OSTBC with 256-6D at 1.5 bpcu in a Rayleigh slow fading channel with $f_d = 10^{-5}$. We observe that the QOSTBC outperforms the OSTBC with 256-6D. Indeed, the constellations optimized using bounds on the performance of uncoded systems, despite benefiting from a high average mutual information at high SNR, do not work well when concatenated with FEC codes due to their low average mutual information at low-to-moderate SNRs, and a bad constellation structure for the outer FEC code.

However, in Fig. 6.20, we showed an example of use cases of the optimized constellations when concatenated with polar codes. Indeed, multidimensional constellations, introduced in this chapter, can be constructed for especial spectral efficiencies when the construction of other STBCs is difficult, e.g., at low spectral efficiencies. These constellations in conjugation with polar codes can construct power efficient schemes.

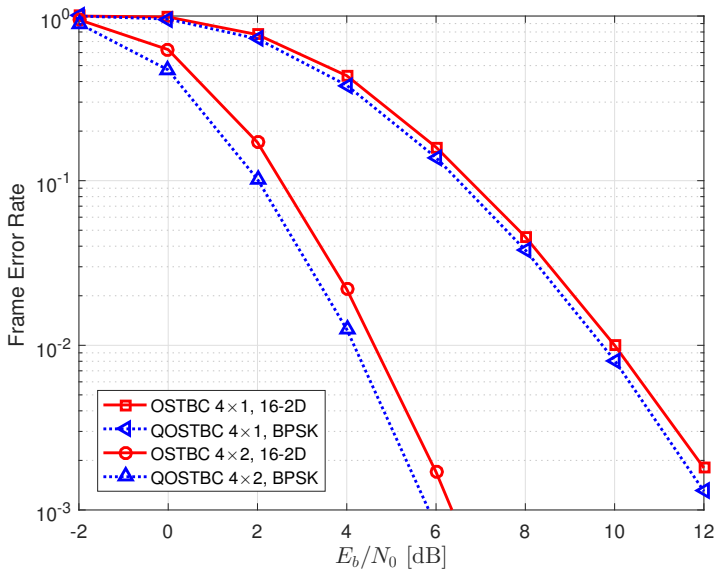



**Figure 6.21:** FER comparison of the MLPCM scheme constructed using QOSTBC with QPSK and OSTBC with a 256-6D constellation in a Rayleigh slow fading channel.

As an example, while $2^{2\bar{B}}$-QAM constellations seem to perform well with FEC codes in comparison to other power efficient 16-2D constellations, $2^{2\bar{B}+1}$-QAM constellations do not work well with FEC codes and the optimized constellations, e.g., for $m = 10$ can perform better in both AWGN and Rayleigh fading channels.

## 6.5 Conclusion

In this chapter, two bounds on the performance of the SER and BER of multidimensional constellations in a Nakagami-$m$ fading channel at high SNR are derived. These bounds are used to obtain the constellations that minimize the SER and BER. The convexity of the SER and BER upper bounds is proven for 1D constellations. As the result of optimization, SER-minimizing 1D, 2D and multidimensional constellations overcome the best-known constellations in SISO configuration. The BER-minimizing constellations optimized using the BER bound outperform the constellations found

based on the independent optimization of shape and the bit-to-symbol mapping. In addition, it is shown that adapting the constellations based on the channel parameter $m$ can improve the performance.

The multidimensional constellations were also optimized for the OSTBC and it is observed that these constellations can improve the BLER of the OSTBC in comparison to regular 2D QAM constellations. The OSTBC with multidimensional constellations works well for low-to-moderate spectral efficiencies where all degrees of freedom of algebraic codes cannot be fully exploited. Even though the non-orthogonal algebraic codes may provide better performance than orthogonal STBCs at high spectral efficiencies, the optimized constellations provide a trade-off between decoding complexity and performance. Furthermore, the scheme outperforms QOSTBCs but this improvement is achieved at the expense of higher ML decoding complexity. Finally, the proposed generalized scheme can provide a complexity-performance trade-off for OSTBCs with multidimensional constellations. In this chapter, the performance advantage of optimized constellations in presence of polar codes is demonstrated. In particular, they can provide better performance at spectral efficiencies, in which the construction of regular multidimensional constellations or STBCs is difficult. One of the shortcomings of multidimensional constellations, that makes the optimization quite difficult, is their large number of variables. Furthermore, the optimization of multidimensional constellations based on the error bounds does not guarantee high mutual information or low outage probability at low-to-moderate SNRs.

## 6.6 Appendix A

### 6.6.1 Derivation of (6.14): SER Union Bound for a Nakagami-$m$ Channel

By using the Chernoff bound, (6.13) is upper bounded as

$$\begin{aligned} \mathrm{P}(\mathbf{s} \to \hat{\mathbf{s}} \mid \mathbf{H}) &\leq \exp\left(-\left(\frac{\sum_{n_r=1}^{N_r}\sum_{n_t=1}^{N_t}|h_{n_t,n_r}|^2}{4N_0}\right)\sum_{\kappa=1}^{K}|s^{(\kappa)}-\hat{s}^{(\kappa)}|^2\right) \\ &= \prod_{j=1}^{N_r}\prod_{i=1}^{N_t}\exp\left(-\left(\frac{|h_{n_t,n_r}|^2}{4N_0}\right)\sum_{\kappa=1}^{K}|s^{(\kappa)}-\hat{s}^{(\kappa)}|^2\right). \end{aligned} \tag{6.29}$$

By considering the distribution of $|h_{n_t,n_r}|^2$ for Nakagami-$m$ fading, the PEP can be upper bounded further as

$$\mathrm{P}(\mathbf{s} \to \hat{\mathbf{s}}) \leq \frac{4^{mN_tN_r}}{\prod_{n_r=1}^{N_r}\prod_{n_t=1}^{N_t}\left(\bar{\gamma}_{n_t,n_r}\frac{\sum_{\kappa=1}^{K_C}|s^{(\kappa)}-\hat{s}^{(\kappa)}|^2}{m}+4\right)^m}. \tag{6.30}$$

From (6.30), a union bound on the SER of OSTBC with multidimensional constellations can be derived as (6.14).

### 6.6.2 Derivation of (6.17): The Second SER Union Bound for a Nakagami-$m$ Channel

At high SNR where $\bar{\gamma}_{n_t,n_r}/m >> 4$, the PEP in (6.30) can in turn can be bounded by

$$\mathrm{P}(\mathbf{s} \to \hat{\mathbf{s}}) \leq \frac{(4m)^{mN_tN_r}}{\left(\prod_{n_r=1}^{N_r}\prod_{n_t=1}^{N_t}\bar{\gamma}_{n_t,n_r}\right)\left(\sum_{\kappa=1}^{K_C}|s^{(\kappa)}-\hat{s}^{(\kappa)}|^2\right)^{mN_tN_r}}. \tag{6.31}$$

From (6.31), the corresponding union bound on the SER can be derived as (6.17).

### 6.6.3 Derivation of (6.21): SER Union Bound for an AWGN Channel

To prove (6.21), starting from (6.29) and by considering $|h_{n_t,n_r}|^2 = 1$ for the AWGN channel, the PEP upper bound is written as

$$\mathrm{P}(\mathbf{s} \to \hat{\mathbf{s}}) \leq \exp\left(-\left(\frac{N_r N_t}{4N_0}\right)\sum_{\kappa=1}^{K_C}|s^{(\kappa)} - \hat{s}^{(\kappa)}|^2\right). \tag{6.32}$$

Finally, by considering the union bound, (6.21) can be derived.

## 6.7 Appendix B

### 6.7.1 Proof of Convexity of (6.14) for 1D Constellations

For simplicity let us assume $\bar{\gamma}_{n_t,n_r}$ is the same for all paths, i.e., $\bar{\gamma}' = \bar{\gamma}_{n_t,n_r}$. For 1D constellations, (6.30) can be written as

$$\mathrm{P}(\mathbf{s} \to \hat{\mathbf{s}}) \leq \frac{C_B''}{\left(\frac{\bar{\gamma}'}{m}(s-\hat{s})^2 + 4\right)^{mN_tN_r}}, \tag{6.33}$$

where $C_B'' = 4^{mN_tN_r}$. Setting $\hat{\delta}_s = (s - \hat{s})$ and $x = C_B''(\frac{\bar{\gamma}'}{m}\hat{\delta}_s^2 + 4)^{-mN_tN_r}$, we take the second derivative of $x$ to examine the convexity. It can be given as

$$\frac{d^2x}{d\hat{\delta}_s^2} = 2\bar{\gamma}' N_t N_r C'' \frac{(2mN_tN_r + 1)\hat{\delta}_s^2\bar{\gamma}'/m - 4}{(\frac{\bar{\gamma}'}{m}\hat{\delta}_s^2 + 4)^{2+mN_tN_r}}. \tag{6.34}$$

For convexity of (6.30), $d^2x/d\hat{\delta}_s^2$ should be positive. This only happens if the following condition is satisfied:

$$x' \coloneqq \frac{2}{\sqrt{\frac{\bar{\gamma}'}{m}(2mN_tN_r + 1)}} < |\hat{\delta}_s|. \tag{6.35}$$

Therefore, the minimum distance of the constellation points should be greater than $x'$. For OSTBCs, the total average SNR of each received matrix $\mathbf{G}$ is simplified as $\bar{\gamma} = E[\sum_{n_t}\sum_{n_r}|h_{n_t,n_r}|^2]/N_0 = N_tN_r/N_0$, which results in $\bar{\gamma} = N_tN_r\bar{\gamma}'$. Therefore, $x'$ can be rewritten in terms of the total average SNR as (6.16). Given the convexity condition in (6.35), the upper bound on the PEP in (6.33) is convex on mutually excluded sets since either $\hat{\delta}_s > x'$ or $\hat{\delta}_s < -x'$ for each pair of points. Because the sum of convex functions preserves the convexity [142], the union bound on the PEP in (6.17) is also convex in the convexity regions of (6.33) given in (6.15). The more general case, where $\bar{\gamma}_{n_t,n_r}$ is different for each path, can be proven by considering the log-convexity of (6.30) for 1D constellations. The log-convexity of pairwise error probabilities is also used in [128] in the case of constellation design for a relay channel.

### 6.7.2 Proof of Convexity of (6.17) for 1D Constellations

For 1D constellations, (6.31) can be written as

$$\mathrm{P}(\mathbf{s}\to\hat{\mathbf{s}}) \le \frac{C_B''}{(s-\hat{s})^{2mN_tN_r}}, \tag{6.36}$$

where $C_B'' = (4m)^{mN_tN_r}/\bar{\gamma}$. We set $y = C_B''(s-\hat{s})^{-2mN_tN_r}$. Hence $d^2x/d\hat{\delta}_s^2$ can be given as

$$\frac{d^2x}{d\hat{\delta}_s^2} = 2mN_tN_rC_B''(2mN_tN_r+1)\delta^{-2(mN_tN_r+1)}. \tag{6.37}$$

For convexity of (6.37), it is sufficient that $\hat{\delta}_s > 0$. First, we explain why $\hat{\delta}_s$ can be non-negative. Without loss of generality, we can assume a specific ordering for 1D constellations, since symbol labelling is not effective on the SER. Thus, we can set $s_1 \le s_2 \le \ldots \le s_{2^B}$. Therefore, we can assume $\hat{\delta}_s$ is always non-negative so the convexity condition is limited to $\hat{\delta}_s \neq 0$. This observation is consistent with the

asymptotic behavior of (6.16) since, as $\bar{\gamma}$ tends to infinity, $x$ tends to zero. Therefore, PEP in (6.36) is a convex function on mutually excluded sets corresponding to $\hat{\delta}_s > 0$ or $\hat{\delta}_s < 0$. The rest of the proof is the same as the last part in Appendix 6.7.1.

## 6.8 Appendix C: A Proof for the Simplified Decoder of OSTBCs

Here, a proof for the simplified decoder of OSTBCs is presented. Due to the algebraic complexity, we show the proof only for the case of two transmit and one receive antennas. Proving all other cases is strait forward using the same steps. We start with the signal definition and the ML decoder:

$$\begin{aligned} y_1 &= h_{1,1}s_1 + h_{2,1}s_2 + w_1, \\ y_2 &= h_{1,1}s_2^* - h_{2,1}s_1^* + w_2, \end{aligned} \tag{6.38}$$

$$\mathrm{P}(y_1, y_2|s_1, s_2) = \frac{1}{\pi N_0} e^{-\frac{1}{N_0}(|y_1 - h_{1,1}s_1 - h_{2,1}s_2|^2 + |y_2 - h_{1,1}s_2^* + h_{2,1}s_1^*|^2)}. \tag{6.39}$$

After some extensions, (6.39) can be written as

$$\begin{aligned} \mathrm{P}(y_1, y_2|s_1, s_2) = \frac{1}{\pi N_0} & e^{-\frac{1}{N_0}\left(|y_1|^2 + |y_2|^2 - \Psi_1^* s_1 - \Psi_1 s_1^* - \Psi_2^* s_2 - \Psi_2 s_2^*\right)} \\ & .e^{-\frac{1}{N_0}\left(|h_{1,1}|^2|s_1|^2 + h_{1,1}h_{2,1}^* s_1 s_2^* + h_{1,1}^* h_{2,1} s_1^* s_2 + |h_{2,1}|^2|s_2|^2\right)} \\ & .e^{-\frac{1}{N_0}\left(|h_{2,1}|^2|s_1|^2 - h_{1,1}h_{2,1}^* s_1 s_2^* - h_{1,1}^* h_{2,1} s_1^* s_2 + |h_{1,1}|^2|s_2|^2\right)}, \end{aligned} \tag{6.40}$$

where

$$\begin{aligned} \Psi_1 &= h_{1,1}^* y_1 - h_{2,1} y_2^*, \\ \Psi_2 &= h_{1,1} y_2^* + h_{2,1}^* y_1. \end{aligned} \tag{6.41}$$

Finally, (6.40) can be simplified as

$$\mathrm{P}(y_1, y_2|s_1, s_2) = \frac{1}{\pi N_0} e^{-\frac{1}{N_0}\left(|y_1|^2+|y_2|^2-\frac{|\Psi_1|^2}{\bar{K}}-\frac{|\Psi_2|^2}{\bar{K}}\right)} . e^{-\frac{1}{\bar{K}N_0}|\Psi_1-\bar{K}s_1|^2} . e^{-\frac{1}{\bar{K}N_0}|\Psi_2-\bar{K}s_2|^2}, \quad (6.42)$$

where

$$\bar{K} = \sum_{n_t}\sum_{n_r}|h_{n_t,n_r}|^2. \quad (6.43)$$

The first exponential term in (6.42) remains the same for all constellation points and therefore, it can be neglected in decoding. The second and the third exponential terms show that the ML decoding of the first and the second symbols of a STBC is separable. The same approach can be used to simplify the decoder for larger STBCs.

# Chapter 7

# Coded Space-Time Signal Design in Slow Fading Broadcast Channels

Slow fading broadcast channels can model a wide range of wireless applications in which the channel fading coefficients change much more slower than the transmission rate (e.g., the control channel for moving user front-ends). When the wireless application is delay-tolerant, multiple realizations of the fading coefficients can be employed to observe an ergodic channel. However, many applications in slow fading broadcast channels are delay-sensitive and due to the difference of channel state information (CSI) experienced by different users, the CSIT is not available or cannot be used. Therefore, the observed channels are non-ergodic. Consequently, for modulation and coding across these channels, the high throughput adaptive modulation and coding schemes cannot be employed. Instead, the signal should provide on average reliable throughput for all users. To achieve reliable throughput, the signal can be designed to minimize the average FER at an average SNR for given channel statistics (e.g., Rayleigh fading). Alternatively, to design reliable schemes, the outage probability can be minimized since it is a lower bound on the FER of the system in non-ergodic fading broadcast channels [56, 156, references therein].

MIMO systems can exploit spatial diversity and multiplexing to minimize the

outage probability in fading channels. STBCs are a class of low-complexity MIMO schemes that can achieve low outage probability without CSIT. Thus, STBCs are a reasonable solution for communicating over slow fading broadcast channels. OSTBCs, introduced by Alamouti in [110] and by Tarokh, Jafarkhani and Calderbank in [133], benefit from low decoding complexity and can use the full spatial diversity. However, they cannot achieve any coding gain and suffer from a rate loss as the number of antennas grows. To overcome these limitations, QSTBCs [135], super-orthogonal space-time trellis codes [157], algebraic codes [35, 158], [138, references therein], and space-time super-modulations [159] have been proposed.

Tarokh et al., in [33], introduced the rank and determinant criteria as two useful measures for designing STBCs at high SNRs. Even though most STBCs are designed on the basis of these criteria, they do not guarantee good performance at low-to-moderate SNRs [55]. Instead, STBCs can be optimized by minimizing the outage probability at a given SNR.

Typically, since MLC/MSD achieves the channel capacity, the symbol-wise average mutual information is maximized for signal design assuming that capacity-achieving FEC codes are employed [53]. Similarly, as we will show in this chapter, the symbol-wise outage probability can be minimized to achieve reliable STBCs for non-ergodic channels. To estimate the outage with low-complexity, bounds on the outage probability may be employed [57]. However, deriving tight outage bounds for the general form of STBCs presents difficulties. Alternatively, as we will show in Section 7.4, an upper bound on the outage probability of the general form of STBCs can be derived by substituting the cutoff rate [8, 160, 161] instead of the mutual information. Although this bound models the diversity and can be used to suggest a general form for the signal, it is not tight enough to achieve good coding gain. Thus, due to the lack of a tight closed form for the outage probability, we propose a relatively simple optimization method to minimize the outage directly.

As we will show in this chapter, optimization on the basis of the outage probability is especially useful in designing classes of STBCs that cannot be designed based on the rank and determinant criteria. One of these classes is SBCs. SBCs, also known as space only-codes, are a special case of STBCs in which the MIMO signal is only distributed on antennas and does not use the time. In this case, a binary FEC outer code with low-to-moderate code rates can compensate the coding gain that is achieved using time in STBCs, while for high rate FEC outer codes, employing STBCs may still provide substantially better performance. However, STBCs other than the OSTBCs, typically suffer from high LLR estimation complexity in comparison to SBCs. Even though this corresponds to a performance-complexity trade-off, the low-complexity of SBCs motives the use of them for many applications.

Increasing the diversity corresponds to decreasing the probability of bad scaling and rotation in a fading channel. To increase the robustness of the signal against the worse case rotation, constraints on the signal structure can be proposed based on the approximate universality criterion for STBCs [162]. However, this can highly increase the signal dimensionality and the LLR estimation complexity. In a FEC coded system, the codeword can be averaged out many space-time blocks. Therefore, to decrease the effect of the worse case rotation, time-varying SBCs (TVSBCs) with a random rotation in each space-time block rotation can be employed [37, 38].

For designing SPM, the direct evaluation of the open form metrics on the performance of MSD, such as the sum of binary channel cutoff rates [58], is difficult. Instead, typically channel dependent pairwise metrics are used to design set-partitioning [163]. One relevant pairwise metric for slow fading channels is the pairwise outage probability [164]. Fortunately, by employing the cutoff rate based outage bound, a closed form upper bound on the pairwise outage probability can be derived. As we will show in the results, it can improve the performance of the coded-modulation scheme.

The signal design based on the outage probability can improve the performance

especially for long FEC codes [57]. However, since it does not consider the structure of the FEC codes and decoders, it is not the best metric for designing signals used with short to moderate length codes. Therefore, novel objective functions based on the structure of FEC codes, decoders, and the bit-to-symbol mapping may improve the performance of the joint coded-modulation scheme.

To achieve the best match between polar codes and STBCs, the parameters of the STBC can be adapted to the structure of the polar code and the corresponding SCD. In this chapter, we propose a method for optimizing multilevel polar codes and STBCs jointly. To this end, we change the parameters of the STBC to create a new space-time code, find a good SPM and optimize the polar code for the new STBC. We repeat these steps consecutively until the best match is achieved. Here, given a specific code rate and average SNR, the design objective of the concatenated polar code and STBC is to minimize the FER.

In this chapter, we aim to enhance the performance of the concatenation of STBCs and multilevel polar codes for slow fading broadcast channels. Thus, we optimize STBCs, low-complexity SBCs [74, 81], and TVSBCs [37] using the outage probability and the joint optimization method. We derive an upper bound on the outage probability and the pairwise outage probability to design SPM. We also derive the outage optimality condition for slow fading channels. In particular, we show that for MSD, outage probabilities of all levels are optimally equal. Based on this observation, we propose an outage rule to determine component code rates of a MLC/MSD scheme. Here, all methods are only used for the offline optimization of small MIMO systems.

The rest of the chapter is organized as follows: in Section 7.1, the system model is defined; in Section 7.2, the STBC design methods and codes used in this chapter are reviewed; in Section 7.3, the STBC design by minimizing the outage probability is described; in Section 7.4, the bounds on the outage probability are derived; in Section 7.5, the design elements of MLPCM, including the outage rule for determining

component code rates, the labelling algorithm, and the MLPCM design procedure are discussed; in Section 7.6, the joint optimization of polar codes and STBCs is explained; in Section 7.7, the numerical results are presented; and in Section 7.8, the conclusions are provided.

## 7.1 System Model

The system model is defined in Chapter 2 in which the CSIT is not available.

## 7.2 Review of STBC Design Methods

In this section, we review the structure and the design methods of STBCs, SBCs, and TVSBCs used in this chapter.

### 7.2.1 Space-Time Block Codes

The rank and determinant criteria are extremely effective in minimizing the union bound of the PEP of STBCs at high SNRs. The rank of the pairwise difference matrix, $\Delta$, as a measure of the diversity, and the determinant of $\Delta\Delta^{\mathcal{H}}$, as a measure of the coding gain for STBCs have been widely used for the design of codes. One of the well-known STBCs designed with this method, capable of achieving full diversity and the highest coding gain for a $2 \times 2$ antenna configuration, is the Golden code introduced by Belfiore et al. in [35]. The structure of the Golden code, herein referred to as Matrix $\mathbf{G}_6$, is given by

$$\mathbf{G}_6 = \frac{1}{\sqrt{5}} \begin{bmatrix} \alpha(s_1 + s_2\theta) & \alpha(s_3 + s_4\theta) \\ \jmath\bar{\alpha}(s_3 + s_4\bar{\theta}) & \bar{\alpha}(s_1 + s_2\bar{\theta}) \end{bmatrix}, \tag{7.1}$$

where $s_1$, $s_2$, $s_3$, and $s_4$ are symbols of the code chosen from a constellation with cardinality $|2^{\hat{B}}|$, $\theta = (1+\sqrt{5})/2$, $\bar{\theta} = 1-\theta$, $\alpha = 1+\jmath(1-\theta)$, and $\bar{\alpha} = 1+\jmath(1-\bar{\theta})$, and $\jmath = \sqrt{-1}$. The Golden code suffers from high decoding complexity. Later, Sezigner and Sari in [158] introduced a $2\times 2$ STBC to reduce the decoding complexity of the Golden code. This code, herein referred to as Matrix $\mathbf{G}_7$ and can be written as

$$\mathbf{G}_7 = \begin{bmatrix} \alpha_1 s_1 + \alpha_2 s_3 & \alpha_1 s_2 + \alpha_2 s_4 \\ -\beta_1 s_2^* - \beta_2 s_4^* & \beta_1 s_1^* + \beta_2 s_3^* \end{bmatrix}, \tag{7.2}$$

where $\alpha_1 = \beta_1 = 1/\sqrt{2}$, $\alpha_2 = 1/\sqrt{2}e^{\jmath\phi}$, $\beta_2 = -\jmath\alpha_2$ and $*$ denotes the conjugate. Using a numerical search, $\phi = 114.29°$ is found to maximize the minimum determinant for QPSK [158]. As shown in [158], $\mathbf{G}_7$ performs only slightly worse than $\mathbf{G}_6$.

For capacity-achieving codes, the constellation or STBC that maximizes the average mutual information is one of the best modulations for achieving a performance close to capacity [53]. Therefore, maximizing the average mutual information can be used as a SNR-dependent metric for design and analysis of STBCs [165]. The mutual information for a given realization of $\mathbf{H}$ can be written as

$$\mathcal{I}(\mathbf{Y};\mathbf{G}|\bar{\gamma},\mathbf{H}) = \sum_i \mathrm{P}(\mathbf{G}_i) E_{\mathbf{Y}}\Big[\log_2\Big(\frac{\mathrm{P}(\mathbf{Y}|\mathbf{G}_i,\mathbf{H},\bar{\gamma})}{\sum_j \mathrm{P}(\mathbf{G}_j)\mathrm{P}(\mathbf{Y}|\mathbf{G}_j,\mathbf{H},\bar{\gamma})}\Big)\Big]. \tag{7.3}$$

For notational convenience, $\mathcal{I}(\mathbf{Y};\mathbf{G}|\bar{\gamma},\mathbf{H})$ is denoted as $\mathcal{I}$ hereinafter. The average mutual information can be estimated by taking the expectation over $\mathbf{H}$ as $\mathcal{I}(\mathbf{Y};\mathbf{G}|\bar{\gamma}) = E_{\mathbf{H}}[\mathcal{I}]$.

In [55], to achieve good performance at low SNRs, the Matrix $\mathbf{G}_7$ code is optimized by maximizing the mutual information at low SNRs. As a result, the parameters of the Matrix $\mathbf{G}_6$ code for low SNRs are different from the code designed based on the rank and determinant criteria. For example, for a wide range of low-to-moderate SNRs and by employing QPSK, $\phi$ is found to be 135° for $N_r = 2$.

### 7.2.2 Space Block Codes and Time-varying Space Block Codes

As one of the simplest forms of STBCs, OSTBCs benefit from low-complexity decoding. However, most high-performance STBCs have unnecessarily high decoding complexity. Indeed, a strong binary FEC outer code can compensate for most of the coding gain of the STBC that is achieved by timely expansion of the signal. Therefore, to reduce the complexity of STBC detection, the code can be expended only across the antennas (i.e. $T = 1$) to increase the transmission rate compared to single input multiple output systems. This class of STBCs is herein referred to as SBCs. In this section, we review the structure and design method of SBCs. The simplest form of SBCs is the vector channel symbol introduced by Hochwald and ten Brink in [74]. This code can be written as

$$\mathbf{G}_8 = [s_1 \quad s_2 \quad \ldots \quad s_{N_t}] \tag{7.4}$$

where $s_{n_t}$ is chosen from a QAM constellation. Instead of using available constellations in the structure of SBCs, parameterized SBCs can be optimized according to channel statistics. A parameterized SBC introduced in [166], herein referred to as $\mathbf{G}_9$, can be written as

$$\mathbf{G}_9 = [\alpha_1 s_1 + \beta_1 s_2 \quad \alpha_2 s_1 + \beta_2 s_2]. \tag{7.5}$$

Using $\mathbf{G}_9$, the received vector can be given as $\mathbf{y} = \mathbf{sVH}+\mathbf{w}$ where $\mathbf{s} = [s_1 \quad s_2]$ and $\mathbf{V}$ is given as

$$\mathbf{V} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix}. \tag{7.6}$$

To achieve the maximum diversity for $\mathbf{G}_9$, **sVH** must be of full rank. Thus, the determinant of **H** and **V** should be non-zero. We know that $|\mathbf{H}| \neq 0$ since any matrix with random variable elements will be full rank with probability one [56].

In contrast, **V** should be adjusted to have a non-zero determinant. This results in $\alpha_1\beta_2 \neq \alpha_2\beta_1$ which must be used when designing $\mathbf{G}_9$. In $\mathbf{G}_9$ only two symbols are employed. By increasing the number of symbols, the number of degrees of freedom for the optimization increases. To this end, we propose to use four different symbols in the structure of a SBC. The corresponding SBC structure can be given as

$$\mathbf{G}_{10} = \begin{bmatrix} \alpha_1 s_1 + \beta_1 s_2 & \alpha_2 s_3 + \beta_2 s_4 \end{bmatrix}. \tag{7.7}$$

The matrix **V** for $\mathbf{G}_{10}$ can be written as

$$\mathbf{V} = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 \\ 0 & 0 & \alpha_2 & \beta_2 \end{bmatrix}^{\mathcal{T}}, \tag{7.8}$$

where $\mathcal{T}$ is the transpose operator. In this case, as soon as all coefficients are non-zero, the matrix **V** is of full rank.

In the next sections, we optimize the SBCs using different metrics. All these codes are designed on the basis of the following principles. First, to limit the emitted power of each antenna, we assume $|\alpha_1|^2 + |\beta_1|^2 = |\alpha_2|^2 + |\beta_2|^2$. Next, to limit the search space, we set $\alpha_2 = \alpha_1$ and $\measuredangle\alpha_1 = 0°$. To satisfy $\alpha_1\beta_2 \neq \alpha_2\beta_1$ for $\mathbf{G}_9$, we can set $\beta_2 = \jmath\beta_1$. Without loss of generality, the same condition can be used for $\mathbf{G}_{10}$. Therefore, the phase of $\beta_1$, $\varphi = \measuredangle\beta_1$, and the ratio of magnitudes $|\beta_1|/|\alpha_1|$ should be optimized. To keep the total power constant, a power constraint is considered which can be given as

$$\frac{1}{N_t T} \sum_{n_t} \sum_{t} |g_{n_t}^t|^2 = 1. \tag{7.9}$$

SBCs with higher number of transmit antennas can substantially enhance the performance. In this chapter, we also optimize the extension of $\mathbf{G}_{10}$ for $N_t = 3$ given as

$$\mathbf{G}_{11} = [\alpha_1 s_1 + \beta_1 s_2 \quad \alpha_2 s_3 + \beta_2 s_4 \quad \alpha_3 s_5 + \beta_3 s_6]. \tag{7.10}$$

For $\mathbf{G}_{11}$, to limit the search space, we set $\alpha_1 = \alpha_2 = \alpha_3 = 1$ and optimize other parameters given the power condition $|\alpha_1|^2 + |\beta_1|^2 = |\alpha_2|^2 + |\beta_2|^2 = |\alpha_3|^2 + |\beta_3|^2 = 1/3$.

Introduced by Duyck et al. in [37], TVSBCs can improve the performance of SBCs by providing a wide range of rotations for symbols of each antenna during a codeword transmission. This in turn results in limiting the effect of the worst-case fading rotation and increasing the diversity. For implementing the TVSBC, the corresponding SBC can be multiplied by a $N_t \times N_t$ matrix $\mathbf{A}_R$ which is known at the receiver. The main diagonal elements of $\mathbf{A}_R$ are time-varying random complex numbers with constant magnitudes given as $e^{\jmath\theta_i(t)}$. Therefore, the new TVSBC can be written as $\mathbf{G}(t) = \mathbf{G}\mathbf{A}_R$. Matrix $\mathbf{A}_R$ for the case of $N_t = 2$ can be given as

$$\mathbf{A}_R = \begin{bmatrix} e^{\jmath\theta_1(t)} & 0 \\ 0 & e^{\jmath\theta_2(t)} \end{bmatrix}, \tag{7.11}$$

where $\theta_1(t)$ and $\theta_2(t)$ are time-varying phase values. Since approximately universal codes are already designed to be robust against the worst-case rotation [162], unlike SBCs, they cannot be improved by applying time-varying schemes.

### 7.2.3 Spatial Modulation

SM is a special case of SBCs in which we turn on only one or few transmit antennas depending on the symbol index in a SPM [167, 168]. However, since the number of active transmit antennas is limited, the degrees of freedom to design the signal is limited. This results in degrading the performance compared to SBCs with the same

spectral efficiency. Therefore, it is used when a lower number of radio frequency chains can be employed. In the basic form of SM, during each time slot only one antenna is turned on. Thus, the signal can be written as

$$\mathbf{G}_{12} = [\mathbf{0}_{1:i-1} \quad s_i \quad \mathbf{0}_{i+1:N_t}]. \tag{7.12}$$

Typically, for SM, the bit-to-symbol mapping is designed for the transmit antenna indices and the constellation used on each antenna independently to reduce the complexity of the decoder [167]. The polar coded SM with independent mapping is presented in [51]. However, by joint designing the bit-to-symbol mapping for the transmit antenna indices and the constellation symbols, the system performance can be improved [169]. Indeed, for the latter case, a multidimensional bit-to-symbol mapping is designed for the entire signal.

## 7.3 STBC Design based on the Outage Probability

For non-ergodic channels, outage probability is an achievable bound on the FER of a FEC coded system [56]. Thus, it is a useful metric for the design and analysis of STBCs used with outer channel coding. The outage probability is the probability of the instantaneous mutual information being less than a specific target rate $R_{tot}B$. It can be written as

$$P_{out}(\mathbf{G}, \bar{\gamma}, R_{tot}B) \coloneqq \mathrm{P}(\mathcal{I} < R_{tot}B) = \mathcal{F}(R_{tot}B), \tag{7.13}$$

where $\mathcal{F}(R_{tot}B)$ is the cumulative distribution function (CDF) of $\mathcal{I}$. For the sake of simplicity, the total outage probability of a STBC is denoted by $\epsilon$ hereinafter. Similarly, by substituting the mutual information for the $b^{th}$ address-bit of a specific STBC in (7.13), the level-wise outage probabilities, $\epsilon_i$, can be defined. The outage

probability can be numerically estimated as

$$\epsilon = \int_{\mathbf{H}} \mathbb{1}\big(\mathcal{I}, R_{tot}B\big) f_{\mathbf{H}} d\mathbf{H}, \tag{7.14}$$

where the distribution of $\mathbf{H}$ is given as $f_{\mathbf{H}} = \frac{1}{\pi^{N_t N_r}} \exp(-\mathrm{tr}\{\mathbf{H}^{\mathcal{H}}\mathbf{H}\})$, and $\mathbb{1}(u,v)$ is the unit step function defined as

$$\mathbb{1}(u,v) = \begin{cases} 1 & u < v, \\ 0 & otherwise. \end{cases} \tag{7.15}$$

As a parallel concept, the $\epsilon$-outage capacity for a given outage probability $\epsilon$ is defined as

$$C_{\epsilon} = \mathcal{F}^{-1}(\epsilon). \tag{7.16}$$

Similarly, $C_{\epsilon_i,i}$ can be defined for different levels of a space-time signal. For a given $\epsilon_i$, $C_{\epsilon_i,i}$ can be approximated numerically by estimating the outage probability for a large number of target rate values between 0 and 1 and choosing the one corresponding to the target $\epsilon_i$.

The outage probability MLPCM and BIPCM constructed using the bit-to-symbol mapping methods proposed in Chapter 3 with $\mathbf{G}_6$ and $\mathbf{G}_8$ are compared in Fig 7.1 for a variety of spectral efficiencies. Time-varying schemes are abbreviated as TV in this figure. As the immediate observation, the outage probability for BICM scheme is worse than MLC/MSD. For BICM, time-varying scheme substantially improves the outage probability. Once a bad rotation in a slow fading channel occurs, time-varying schemes by rotating the signal in each channel-use can make the detection of some symbols possible since the codeword spans in many channel-uses and some symbols corresponding to each message bit are received correctly. For MLCs, when high rates codes are used, the time-varying scheme can improve the performance due

to overcoming the worst case rotations. The advantage of the time-varying scheme is more observable in the outage probability of BICM, since the bit-to-symbol mapping is very effective on the mutual information of BICM while it is not effective on the mutual information of MLC schemes. Indeed, we cannot design perfect mappings since there is no tight bound to design the mapping for non-orthogonal schemes in a slow fading channel. Instead, the general Euclidean distance has been used to design the mapping.

MLC/MSD used with STBCs optimized by minimizing the outage probability can achieve high performance since, as we show in the following theorem, the outage probability of MLC/MSD approaches the constellation-constrained outage probability for a regular partition. For the definition of the regular partition refer to [106].

**Theorem 5.** *The outage probability of MLC/MSD scheme approaches the constellation-constrained outage probability of the slow fading broadcast channel given a regular partition is used.*

*Proof.* Using the chain rule of the mutual information, we can connect the total mutual information and level-wise mutual informations as follows:

$$\epsilon = \mathrm{P}(\mathcal{I} < R_{tot}B) = \mathrm{P}(\mathcal{I}(\mathbf{Y}; \{\mathbf{G}^1, ..., \mathbf{G}^B\}|\bar{\gamma}, \mathbf{H}) < R_{tot}B) = \mathrm{P}(\sum_{i=1}^{B} \mathcal{I}_i < R_{tot}B), \quad (7.17)$$

where $\mathbf{G}^b$ is the set $\{\mathbf{G} \in \mathcal{X}_b\}$ and $\mathcal{I}_i$ is defined as the level-wise mutual information given by $\mathcal{I}(\mathbf{Y}; \mathbf{G}^b|\{\mathbf{G}^1, ..., \mathbf{G}^{b-1}\}, \bar{\gamma}, \mathbf{H})$. For the last equality, we assumed since the partitions are regular, the level-wise mutual informations are well-defined. Note that since the MLC/MSD scheme achieves the constellation-constraint mutual information for any regular mapping, the mapping does not need to be SNR-adaptive. ■

Note that although constructing regular partitions is difficult, we still can construct good partitions and approach the substantial portion of the outage probability.

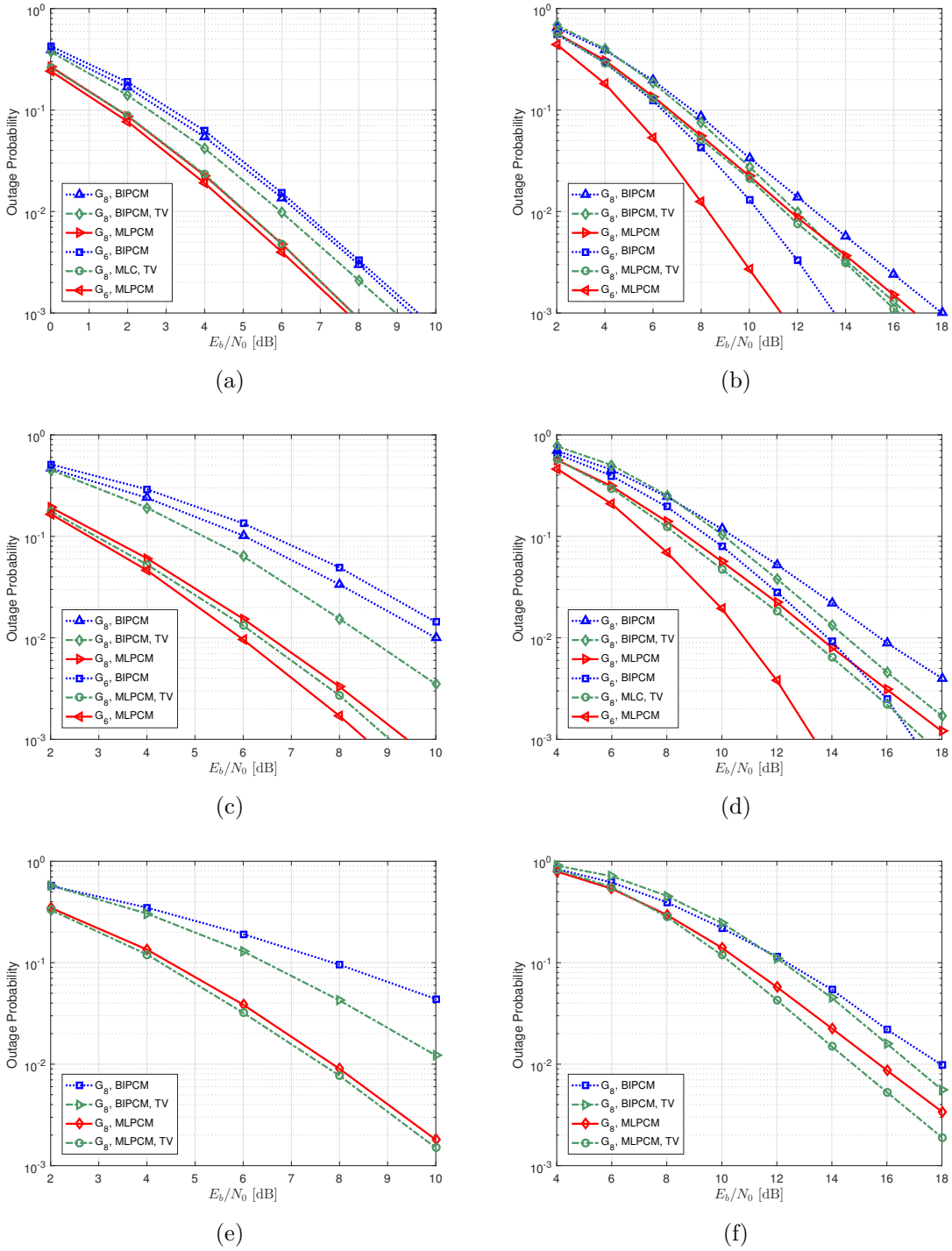



**Figure 7.1:** The outage probability comparison of $\mathbf{G}_6$ and $\mathbf{G}_8$ and TV $\mathbf{G}_8$ with a) $R_{tot}$ = 0.5 and 2 bpcu, b) $R_{tot}$ = 0.9 and 3.6 bpcu, c) $R_{tot}$ = 0.5 and 3 bpcu, d) $R_{tot}$ = 0.9 and 5.4 bpcu, and the comparison of $\mathbf{G}_8$ and TV $\mathbf{G}_8$ with e) $R_{tot}$ = 0.5 and 4 bpcu, and f) $R_{tot}$ = 0.9 and 7.2 bpcu.

We will discuss creating the bit-to-symbol mapping in Section 7.5.2.

To compute the outage probability, due to unavailability of closed-form expressions, (7.14) can be numerically computed or the simulation can be employed. For the numerical optimization, typically the genetic algorithm or the particle swarm optimization (PSO) [170] is employed. In most cases, the genetic algorithm provides a better answer for problems with discrete parameters and the PSO is used when parameters are continuous. In this chapter, since the search space is continuous, the PSO is employed. For the detailed explanation of steps, see [171] and the references therein. For the faster convergence of the algorithm, we modified the PSO according to the following principles. As the FER decreases, the number of realizations of $\mathbf{H}$ to achieve an accurate estimation of the mutual information or the outage probability should be increased. When the PSO starts, a small number of realizations of $\mathbf{H}$ is enough to achieve a coarse estimation of the outage probability for the initial population since they are chosen randomly and thus are most likely far from a local or global minima. Therefore, we increase the number of realizations of $\mathbf{H}$ to estimate the mutual information or the outage as the best function value is lowered.

The modified PSO optimization is formalized in Algorithm 8. The algorithm iteratively updates the position of particles using the randomly generated vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ and constant parameters $\omega_1$ and $\omega_2$. Vector $\mathbf{u}_1$ and parameter $\omega_1$ try to return the $i^{th}$ particle to its last best position and vector $\mathbf{u}_2$ and parameter $\omega_2$ try to move the particle toward the globally best found solution. The parameter $\hat{N}$ is the number of realizations of $\mathbf{H}$ to determine the outage probability and $\omega_3$ is the corresponding velocity. The algorithm stops when the best solution for the vector of STBC parameters ($\boldsymbol{\zeta}$) does not change for several iterations. Note that to optimize TVSBCs, the TV sequence should also be considered in estimation of outage using (7.14). Thus, when calculating $\mathcal{I}$, we average it out over the TV sequences as well.

**Algorithm 8:** Modified PSO to Minimize the Outage

**Input :** STBC matrix $\mathbf{G}_X$ and $\bar{\gamma}$
**Output:** Optimal STBC parameters $\hat{\boldsymbol{\zeta}}$
**Procedures:**
Update_STBC($\boldsymbol{\zeta}$): Generates all points of $\mathbf{G}_X$ by changing the vector of STBC parameters $\boldsymbol{\zeta}$ with size $N_G$ and normalizes the average transmitted power per STBC matrix.
$[\mathbf{u}_1, \mathbf{u}_2]$=randu($N_G$): Generates uniform random vectors in range $[-1, 1]$ and size $N_G$.

**do**
  **for** $i = 1 : Population\_Size$ **do**
    $[\mathbf{u}_1, \mathbf{u}_2]$=randu($N_G$)
    $\boldsymbol{\zeta}_i' = \boldsymbol{\zeta}_i + \omega_1 \mathbf{u}_1 (\bar{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i) + \omega_2 \mathbf{u}_2 (\hat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)$
    $\mathbf{G}_X$ = Update_STBC($\boldsymbol{\zeta}_i + \boldsymbol{\zeta}_i'$)
    $\epsilon(\boldsymbol{\zeta}_i) = \int_{\mathbf{H}_{\hat{N},i}} \mathbb{1}\big(\mathcal{I}(\mathbf{G}_X), R_{tot} B\big) f_{\mathbf{H}} d\mathbf{H}$
    **if** $\epsilon(\boldsymbol{\zeta}_i) < \epsilon(\bar{\boldsymbol{\zeta}})$ **then**
      $\bar{\boldsymbol{\zeta}}_i = \boldsymbol{\zeta}_i$
      $\hat{N} = \omega_3 \hat{N}$
  $\hat{\boldsymbol{\zeta}}_i = \arg\min_{\boldsymbol{\zeta}_i} \epsilon(\bar{\boldsymbol{\zeta}}_\kappa) \quad \forall \kappa \in Neighborhood$
**while** $\neg$ *Stopping Criteria*
return $\hat{\boldsymbol{\zeta}}$

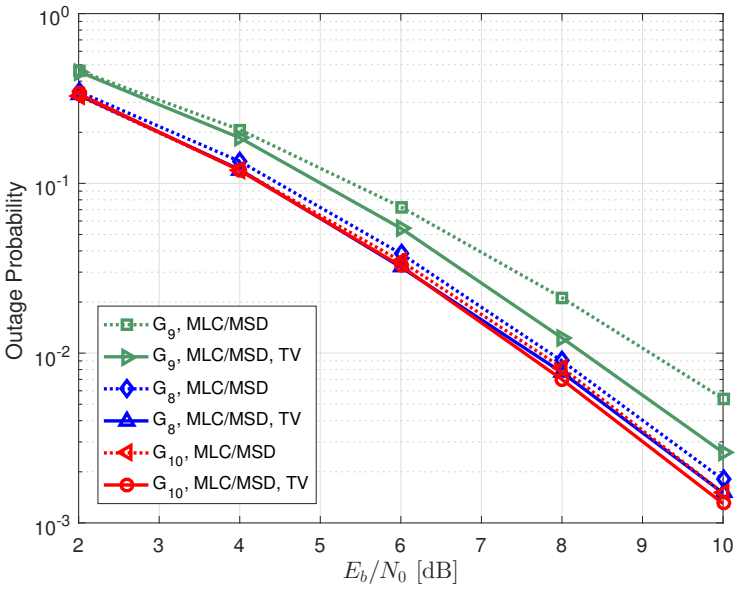
To design STBCs and SBCs using the outage probability, we use the conditions on shaping and power provided in Section 7.2.2 to limit the search space and then we use Algorithm 8 to determine the rest of the parameters. In this chapter, by employing an additional linear search over the SNR [163, Algorithm 2], all STBCs and SBCs are optimized at $\epsilon = 0.01$ for $N_r = N_t$. The optimized values for $\mathbf{G}_7$ for QPSK constellation and $R_{tot} = 1/2$ are given as $\alpha_1 = \beta_1 = 0.314$, $\alpha_2 = 0.067 + 0.381\jmath$, $\beta_2 = -0.070 + 0.384\jmath$. The optimized values of $\alpha_1$, $\beta_1$, and $\varphi$ for $R_{tot} = 1/2$ are as follows: for $\mathbf{G}_9$, 0.5, 0.5, and 330°; for TV $\mathbf{G}_9$, 0.5, 0.5, and 180°; for $\mathbf{G}_{10}$, 0.5, 0.5, and 45°; and for TV $\mathbf{G}_{10}$, 0.5, 0.5, and 45°. The parameters $\alpha_1$, $\beta_1$, and $\varphi$ for $R_{tot} = 9/10$ are as follows: for $\mathbf{G}_9$, 0.5, 0.5, and 270°; for TV $\mathbf{G}_9$, 0.48, 0.52, and 120°; for $\mathbf{G}_{10}$, 0.33, 0.63, and 195°; and

for TV $\mathbf{G}_{10}$, 0.34, 0.62, and 285°.

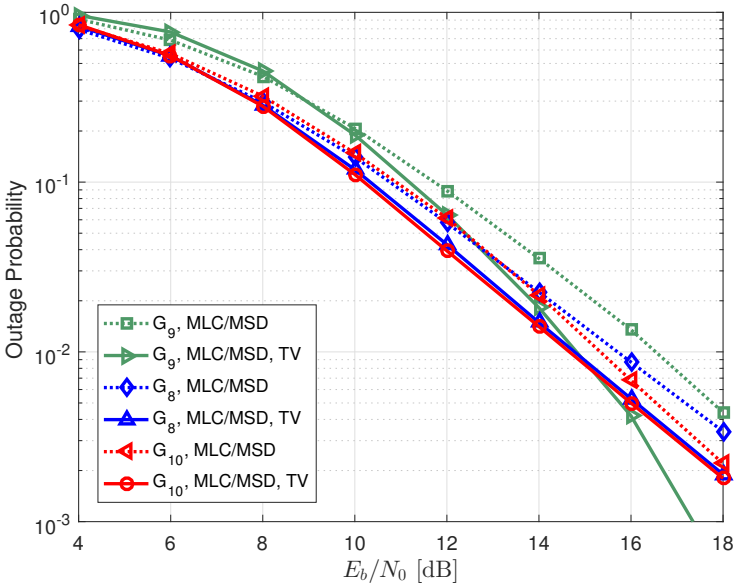
The results of the comparison of outage probabilities of the outage-optimized schemes are shown in Fig. 7.2. For low-to-moderate FEC code rates, TV $\mathbf{G}_{10}$ shows the minimum outage probability but performs very close to $\mathbf{G}_8$. At high rates, for moderate values of the outage probability (e.g., 0.001), $\mathbf{G}_{10}$ shows the least outage probability while for lower values of the outage probability, TV $\mathbf{G}_9$ is better. Indeed, the TV $\mathbf{G}_9$ benefits from a higher diversity, which is useful at high spectral efficiencies, while $\mathbf{G}_{10}$ benefits from more flexibility in optimization, which results in better codes for low-to-moderate spectral efficiencies.

## 7.4 Upper Bound on the Outage Probability

A slow fading channel model and the corresponding signal have slow- and fast-changing parameters. For example, fading coefficients change slowly while AWGN and TV sequences change fast. Thus, we need a class of bounds that can model both effects to achieve good design metrics. Tarokh et al., in [33], analyzed the performance of STBCs using an upper bound on the PEP of STBCs without FEC coding and derived the rank and determinant criteria for the slow fading channel, in which the coefficients of a STBC remains constant during one codeword transmission. However, this method cannot model both slow- and fast-changing parameters. Sezgin and Jorswieck derived a bound on the outage probability of QOSTBCs by finding the achievable fraction of the mutual information in [172]. Although this method of bounding may result in tight bounds, it cannot be generalized to all classes of STBCs due to unavailability of the achievable fraction of the mutual information or the corresponding distribution. Thus, in this section, we derive an upper bound on the outage probability that can model both kinds of parameters.




(a)




(b)

**Figure 7.2:** Outage probability comparison of $\mathbf{G}_8$, $\mathbf{G}_9$, and $\mathbf{G}_{10}$ and their TV variant with a) $R_{tot}$ = 0.5 and 4 bpcu, and b) $R_{tot}$ = 0.9 and 7.2 bpcu, both in a $2 \times 2$ antenna configuration.

Assuming that the cutoff rate is a lower bound on the mutual information, we can derive bounds on the outage probability. The cutoff rate $R_0$ can be written as

$$R_0(\mathbf{G}|\bar{\gamma},\mathbf{H}) = \max_{\mathrm{P}(\mathbf{G}_i)} -\log_2\left\{\sum_{i=1}^{2^B}\sum_{j=1}^{2^B}\hat{\rho}_{i,j}\right\}, \tag{7.18}$$

where $\hat{\rho}_{i,j} = \rho(\mathbf{G}_i,\mathbf{G}_j|\bar{\gamma},\mathbf{H})\mathrm{P}(\mathbf{G}_i)\mathrm{P}(\mathbf{G}_j)$, in which $\rho(\mathbf{G}_i,\mathbf{G}_j|\bar{\gamma},\mathbf{H})$ is the Bhattacharyya coefficient for the pair of STBC points $\mathbf{G}_i$ and $\mathbf{G}_j$. Substituting the cutoff rate, (7.14) can be upper-bounded as

$$\epsilon \leq \mathrm{P}(R_0(\mathbf{G}|\bar{\gamma},\mathbf{H}) < R_{tot}B). \tag{7.19}$$

Note that since the cutoff rate is a lower bound on the mutual information, the rate region of the outage is larger and thus (7.19) is an upper bound on the outage probability. Observe that the cutoff rate part models the fast-changing parameters and the outage probability models the slow-changing parameters. The bound can be written as

$$\mathrm{P}(R_0(\mathbf{G}|\bar{\gamma},\mathbf{H}) < R_{tot}B) = \mathrm{P}\Big(2^{-R_{tot}B} < \sum_{i=1}^{2^B}\sum_{j=1}^{2^B}\hat{\rho}_{i,j}\Big). \tag{7.20}$$

The upper bound in (7.20) in comparison to the outage probability for the Alamouti and the Golden code is depicted in Fig. 7.3 for 2 bpcu.

As we will explain more in Section 7.5.2, to design a SPM, relevant pairwise metrics on the outage probability should be used. In [164], it is shown that the outage probability can be upper bounded by union bounds of pairwise outage probabilities. Therefore, the pairwise outage probability may be employed as a metric to design good SPM. In the rest of this section, using the cutoff rates similar to the upper bound in (7.20), we derive a closed form for the upper bound on the pairwise outage probability (UBPOP). We start with SBCs.

**Proposition 4.** *For SBCs, a UBPOP is given as*

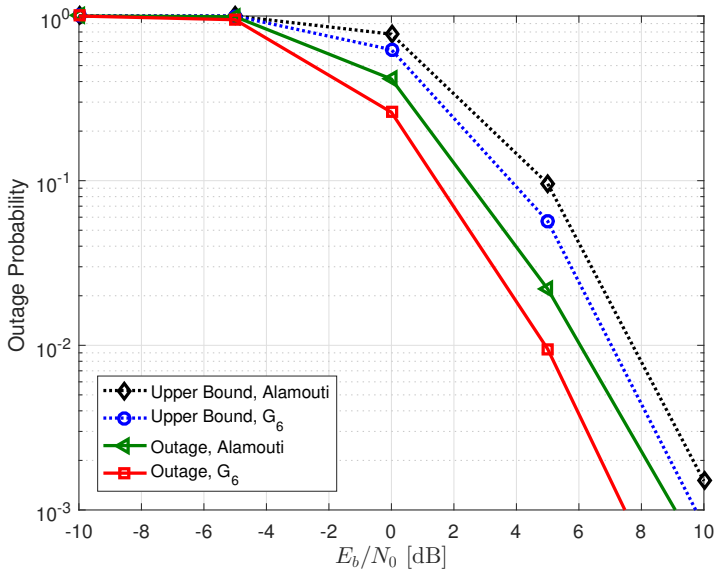



**Figure 7.3:** Comparison of the outage probability and the upper bound in (7.20) for Alamouti and Golden code with 2 bpcu in a $2 \times 2$ antenna configuration.

$$\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j}) = \frac{\gamma_L\big(N_r, -4N_0\|\Delta_{i,j}\|^{-2}\ln(q_{i,j})\big)}{\Gamma(N_r)}, \tag{7.21}$$

*where* $q_{i,j} \geq 0$ *is a constant related to PEP and* $\Delta_{i,j}$ *is the pairwise difference matrix for pairs of* $\boldsymbol{G}_i$ *and* $\boldsymbol{G}_j$ *defined as* $\Delta_{i,j} = \boldsymbol{G}_i - \boldsymbol{G}_j$*, with elements* $\delta_{t,n_t}$ *and* $\gamma_L(\cdot)$ *and* $\Gamma(\cdot)$ *are lower incomplete gamma and gamma functions, respectively.*

*Proof.* In Appendix. 7.9. ∎

Note that $q_{i,j}$ in general should be different for each pair since the distance of pairs in a STBC and correspondingly the effective SNR and their effect on union bounds may be different. However, finding $q_{i,j}$ needs complicated optimizations. Instead, we can find a relatively good value for $q_{i,j}$ as follows. Using (7.18), the pairwise events $\{q_{i,j} < \hat{\rho}_{i,j}\}$ can be written as $\{R_0(\hat{\mathbf{G}}_{i,j}, \bar{\gamma}|\mathbf{H}) < R_{i,j}\}$ where $R_{i,j} = -\log_2(\frac{1}{2} + \frac{1}{2}q_{i,j})$ and $R_0(\hat{\mathbf{G}}_{i,j}, \bar{\gamma}|\mathbf{H})$ is the pairwise cutoff rate. Assuming the average spectral efficiency $R_{tot}$ equals $R_{i,j}$ for all pairs of the STBC, $q_{i,j}$ can be found as $q = 2^{1-R_{tot}} - 1$. Next,

to make the upper bound manageable and find a relevant metric for designing SPM, we can approximate the UBPOP for STBCs as follows:

$$\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j}) \approx \frac{\gamma_L(\frac{\mu_1^2}{\mu_2}, -4N_0\frac{\mu_1}{\mu_2}\ln(q_{i,j}))}{\Gamma(\frac{\mu_1^2}{\mu_2})}, \tag{7.22}$$

where

$$\mu_1 = N_r \sum_t \sum_{n_t} |\delta_{t,n_t}|^2, \quad \mu_2 \approx N_r\Big(\sum_{u=1}^{N_t}\sum_{v=1}^{N_t}|\sum_{n_t}\delta_{u,n_t}\delta^*_{v,n_t}|^2\Big). \tag{7.23}$$

The approximation is derived in Appendix. 7.10.

Note that since we approximate the upper bound in (7.22), it is not anymore an upper bound but it remains a relevant metric for designing SPM. This class of metrics can be used to design the SPM for STBCs. However, for TVSBCs, we need to model TV sequences as well. In this case, we can model the TV sequence using the cutoff rate since it is a fast-changing parameter. An approximation of UBPOP for a $2 \times N_r$ TVSBC is

$$\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j}) \approx 1 - Q\left(\frac{\ln\big(-4N_0\ln(q_{i,j})\big) - \ln\big(\frac{\mu_1}{1+\frac{\mu_2}{\mu_1^2}}\big)}{\sqrt{\ln(1+\frac{\mu_2}{\mu_1^2})}}\right), \tag{7.24}$$

where $Q(.)$ is the Gaussian tail function. The first two moments $\mu_1$ and $\mu_2$ are given as

$$\begin{aligned}
\mu_1 =& N_r\big(|\delta_{1,1}|^2 + |\delta_{1,2}|^2\big) - 2\big(a_1|\delta_{1,1}\delta_{1,2}|E[\Omega] + a_2\big),\\
\mu_2 =& N_r\big(|\delta_{1,1}|^4 + |\delta_{1,2}|^4\big) + 4a_1^2|\delta_{1,1}\delta_{1,2}|^2(N_r - E[\Omega]^2)\\
&- 4a_1N_r\big(|\delta_{1,1}|^2 + |\delta_{1,2}|^2\big)|\delta_{1,1}\delta_{1,2}|(E[|h_{1,n_r}|^2\Omega] - E[\Omega]),
\end{aligned} \tag{7.25}$$

where $\Omega = |\sum_{n_r} h_{1,n_r}h^*_{2,n_r}|$ and coefficients $a_1$ and $a_2$ are defined in (7.36). The approximation is derived in Appendix. 7.11.

## 7.5 Polar Coded-Modulation Design

In this section, we explain different aspects of polar coded-modulation design including determining component code rates and designing the SPM. To design the polar code, the information set should be determined. In this chapter, we use the simulation-based best bit-channels rule proposed explained in Section 4.2.2. In this method, we measure the number of first error events for all bit-channels of all levels and choose bit-channels with lowest error probabilities to determine $\{R_1, ..., R_B\}$.

### 7.5.1 Outage-based Determining Component Code Rates

In this section, we derive a rule based on the outage probability for determining component code rates and compare it with the best bit-channel rule. Two main design rules for determining component code rates of MLC are capacity rule and equal error probability rule for an ergodic channel [106]. When using capacity rule, the component codes rates are chosen based on the constellation constraint subchannel capacity. The capacity rule is only optimal for capacity achieving codes. However, for non-capacity achieving codes, equal error probability rule is typically employed due to the possibility of deriving bounds on the performance of the coded-modulation scheme. For MLC/MSD, as $N \to \infty$, these two rules will converge [1]. For non-ergodic channels, the parallel concepts are the outage rule and the equal error rate rule. The equal error rate rule can be applied similar to the ergodic channel. However, the outage rule needs to be derived.

When adjusting component code rates based on the outage rule, we would like to set $R_i = C_{\epsilon_i,i}$ for the optimal $\epsilon_i$ values that minimize the total outage of the coded-modulation scheme using independent decoding of different levels. Let us denote the total outage probability of sequential decoding of levels as $P_{out}^{SD}$. This is a lower bound on the FER of the coded-modulation scheme.

**Theorem 6.** *$P_{out}^{SD}$ is minimized for a regular SPM if level-wise outage probabilities are equal and for any* **H** *realization, either all levels are in outage or no level is in outage.*

*Proof.* In MSD, an outage event occurs as soon as at least one level is in outage. Thus, we can write $P_{out}^{SD} = \mathrm{P}(\bigcup_i \{\mathcal{I}_i < R_i\})$. Because the probability of the events $\{\mathcal{I}_i < R_i\}$ for all $i = 1, ..., B$ is never less than the maximum of the probabilities of individual events, we have $P_{out}^{SD} \geq \max_i \mathrm{P}(\mathcal{I}_i < R_i)$, with equality holds if the events are fully dependent. Since we assumed for all realizations of **H**, either all levels are in outage or no level is in outage, the events are fully dependent and therefore, we have $P_{out}^{SD} = \max_i \mathrm{P}(\mathcal{I}_i < R_i)$. Now let us assume $C_{\epsilon_i,i}$ is chosen so that the level-wise outage probabilities are equal. Setting $R_1 < C_{\epsilon_1,1}$ results in decreasing $\mathrm{P}(\mathcal{I}_1 < R_1)$. However, since the total $\sum_i R_i$ is constant, we should set $R_2 > C_{\epsilon_2,2}$ that results in increasing $\mathrm{P}(\mathcal{I}_2 < R_2)$. Therefore, the maximum outage probability and consequently $P_{out}^{SD}$ increase. Thus, for optimally we should choose level-wise outage probabilities equal to minimize $P_{out}^{SD}$. ■

Note that in practice when $\epsilon_1 = \epsilon_2, ... = \epsilon_B$ for large FEC codes, we observe that either all levels are in outage or no level is in outage. Therefore, based on the results from Theorem 6, to determine component code rates $R_i$, we should minimize the level-wise outage probability $\epsilon_1 = \epsilon_2 = ... = \epsilon_B$ such that $\sum_i C_{\epsilon,i} = \sum_i R_i = R_{tot}$. Thus, in an attempt to find the code rates, we can estimate $C_{\epsilon_i,i}$s using (7.16) for a given $\epsilon_i$ and check whether $\sum_i C_{\epsilon,i} = R_{tot}$. Note that $\epsilon_i$ should be changed from zero to one with the step length $M$. However, the search space can be substantially limited by starting from $\epsilon_i = \epsilon$. In fact, the total outage of the joint decoding, $\epsilon$, is a lower bound on the outage of different levels since in Theorem 6 we assumed $P_{out}^{SD} = \max_i \mathrm{P}(\mathcal{I}_i < R_i)$ and $P_{out}^{SD} \to \epsilon$ only when $N \to \infty$.

Algorithm 9 presents the outage rule for determining component code rates. In

this algorithm, Function Estimate_Mu(.) estimates $b^{th}$ level mutual information by substituting the probabilities of transmitting zero and one at each level corresponding to (2.5) in (7.3). In each iteration, the level-wise outage capacities $C_{\epsilon_i,i}$ are estimated. If $\sum_i C_{\epsilon_i,i}$ is less than the target spectral efficiency, the threshold $\hat{\epsilon}$ is increased by a factor $M$ and the same steps are repeated. Here, we set $M = 1.05$.

**Algorithm 9:** Outage Rule for Determining Component Code Rates

**Input :** $\mathbf{G}_X$, $\bar{\gamma}$, $R_{tot}$
**Output:** Component code rates $\mathbf{R}$
**Procedures**:
$[\mathbf{I}_{1:B}]$=Estimate_Mu($\mathbf{G}_X$): Estimates the $1^{th}$ to $B^{th}$ level mutual information using (7.3) for $\hat{N}$ realizations of $\mathbf{H}$ and outputs them in $\hat{N} \times B$ matrix $\mathbf{I}_{1:B}$.
Find($\mathbf{u}$,$v$): Outputs $\arg\max_i u_i < v$.

$[\mathbf{I}_{1:B}]$=Estimate_Mu($\mathbf{G}_X$)
$\hat{\epsilon} = \int_{\mathbf{H}} \mathbb{1}\left(\mathcal{I}, R_{tot}B\right) f_{\mathbf{H}} d\mathbf{H}$ ▷ Using (7.14 )
**do**
  **for** *b=1:B* **do** $C_{\epsilon_i,i} = \frac{1}{\hat{N}}$ Find($\mathbf{I}_b$,$\hat{\epsilon}$) ▷ Using (7.16 )
  $\hat{\epsilon} = \hat{\epsilon}M$
**while** $\sum_i C_{\epsilon_i,i} < R_{tot}B$
$\mathbf{R} = [C_{\hat{\epsilon},1}, C_{\hat{\epsilon},2}, ..., C_{\hat{\epsilon},B}]$
return $\mathbf{R}$

Fig. 7.4 shows the comparison of the FERs of the equal FER rule, outage rule, and the simulation-based best bit-channels rule. We observe that the FER of the equal FER rule is 1 dB worse than other two rules and the FER of the best bit channels rule is approximately the same as that of the outage rule. This shows that the simulation-based rule can predict the FERs correctly. For the rest of the chapter, we use the simulation-based rule to determine code rates.

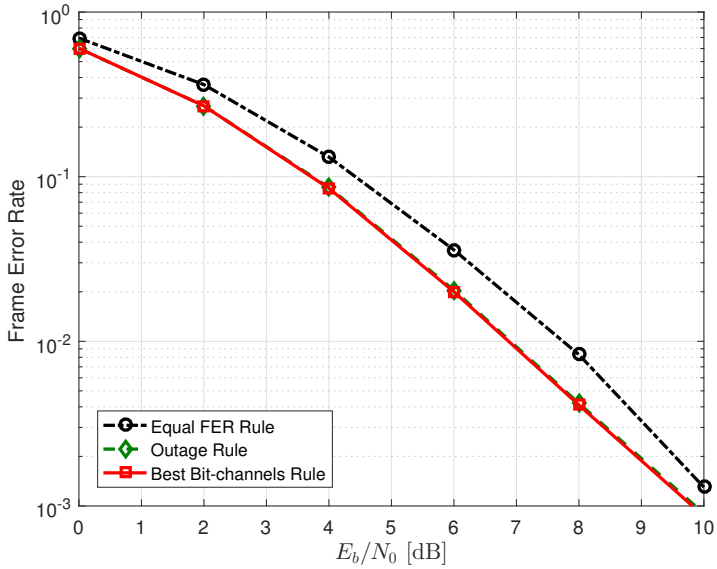



**Figure 7.4:** FER comparison of polar code level rate determining algorithms, all for Alamouti code with 2 bpcu and $N_{tot} = 1024$ in a $2 \times 2$ antenna configuration.

### 7.5.2 Labelling Algorithm

For ergodic channels, an important factor in performance of sequential decoders (e.g., MSD) is the sum of level-wise cutoff rates [58]. For ergodic channels, one of the best mappings to maximize the sum cutoff rates is the Ungerboeck's SPM [59]. MLD/MSD with SPM is also in a close relationship with polar coding [58]. The direct evaluation of the sum cutoff rate for constructing the bit-to-symbol mapping is numerically expensive. Instead, set-partitioning algorithms with relevant channel metrics are used to construct good bit-to-symbol mappings.

In this section, we employ the algorithm reviewed in Section 3.1.3 for creating SPM. In [163] for fast fading channel, the Frobenius norm is used as a measure of distance between subsets. The SPM typically regularizes the component code rates in an incremental order, i.e., $R_1 < R_2 < ... < R_B$. Note that the SPM indeed increases the variability between channel capacities.

For the slow fading channel, based on the result of Theorem 6, we are interested

in equalizing the level-wise outage probabilities. If we assume that the code rates are in an incremental order, we expect that for most fading realizations, the average mutual information of levels is also incremental, i.e., $\mathcal{I}_1 < \mathcal{I}_2 < ... < \mathcal{I}_B$. Thus, we can employ the algorithm in [163] assuming that the appropriate metric is used. If we use approximately universal codes, the corresponding design metric is the determinant criterion when $N_r > N_t$ and the product of smallest $\min(N_t, N_r)$ singular value of $\Delta$ for $N_r < N_t$ [162]. For SBCs, the distance is found to be the Frobenius norm using the PEP analysis according to [33]. Alternatively, UBPOPs provided in Section 7.4 can be used as a measure to construct the mapping. However, it turns out that the use of the determinant criterion and UBPOP for STBCs result in the same SPM. The same result is valid for SBCs when we use the corresponding UBPOP and the Frobenius norm. Although for TVSBCs, the use of UBPOP results in an improvement in comparison to the Frobenius norm.

To overcome the shortcomings of metrics, we can analyze different conditions that may happen for pairwise metrics in a slow fading channel and modify the labelling algorithm in [163]. Remember that the general form of the bounds derived in Section 7.4 is $\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j})$. We expect that as the distance of two pairs increases, $\hat{\rho}_{i,j}$ decreases and correspondingly $-\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j})$ increases. However, due to a variety of reasons, e.g., the effect of pairwise metrics on sum cutoff rate or on the equality of outage probability of different levels, $-\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j})$ may decrease with the distance. But it is unlikely that it happens when the difference in pairwise distances is relatively large. Motived by this explanation, we consider the same value for distances of two pairs of points if the difference of their distances is smaller than a threshold. The threshold is determined in few iterations by constructing the mapping and the polar coded-modulation. In fact, we use the simulation-based design to find the best threshold. In practice, we observe that the threshold remains constant for a

wide range of SNRs. Also while this threshold is substantially effective on unoptimized SBCs, e.g., $\mathbf{G}_8$, it has a negligible effect on optimized SBCs, OSTBCs, and approximately universal STBCs. An algorithm for modification of pairwise metrics is presented in Algorithm 10. In this algorithm, the pairwise metric is computed for all pairs of a STBC and is stored in matrix $\mathbf{D}_G$. Then, the unique values of $\mathbf{D}_G$ are determined and are stored in $\mathbf{s}_{unique}$. Then, by checking each element of $\mathbf{s}_{unique}$, if the difference of $s_{unique,i}$ and $s_{unique,i-1}$ is less than a threshold $M'$ multiplied by one of them, the previous value is assigned to the current one, i.e., $s_{unique,i} = s_{unique,i-1}$. $M'$ varies from zero to one with typical values of 0.1 or 0.2.

**Algorithm 10:** Modification of Pairwise Metrics

**Input :** Pairwise metric matrix $\mathbf{D}_G$ for all pairs of $\mathbf{G}_X$
**Output:** $\mathbf{D}_G$: The modified metric matrix
**Procedures:**
$[\mathbf{s}_{unique}]$=Unique($\mathbf{D}_G$): Determines and sorts all unique values of the pairwise metric matrix, $\mathbf{D}_G$ with elements $d_{i,j}$, for all pairs in vector $\mathbf{s}_u$.

```
[s_unique]=Unique(D_G)
for i = 2 : |s_unique| do
    if |s_unique,i − s_unique,i−1| < M′|s_unique,i| then
        d_i,j = s_unique,i−1, ∀d_i,j == s_unique,i−1
        s_unique,i = s_unique,i−1
return D_G
```

## 7.6 Joint Optimization of FEC Code and STBC

The outage probability is a good criterion for designing the modulation (or STBC) for FEC codes that perform close to the outage. However, when there is a gap between the performance of a FEC code and the outage, the outage probability is not necessarily a good measure. This gap exists for all currently available finite length

FEC codes, even in the presence of powerful decoders, and motivates the research on the joint optimization of STBCs and practical FEC codes.

To optimize concatenated schemes, bounds on the performance of the concatenated FEC code and STBC are needed. For cases such as BICM, if the LLR distribution is nearly Gaussian, general bounds on the performance of the concatenated schemes can be derived [173]. However, for MLC/MSD in which the LLR distribution of each subset within each level is different, deriving any closed form bound on the performance of the system is difficult. Furthermore, optimizing the sum cutoff rate bound is numerically intensive. Instead, the simulation-based design method explained in Section 7.5 can determine the information set and the FER of the limited length MLPCM. In fact, simulation-based design plays the role of a bound on the performance of the system since it can approximate the FER. Thus, using the simulation-based method, joint optimization of limited length FEC codes and STBCs is possible.

In simulation-based polar code design, given a specific code rate and SNR value, the information set for MLPCM is chosen and the FER of the polar coded-modulation is determined [163]. To jointly design the polar code and STBC, the best STBC matched to the polar code structure should be determined. To this end, the optimizer generates a set of parameters of a specific STBC, e.g., $\mathbf{G}_7$. Then, for each combination of parameters of the STBC, the set-partitioning algorithm is applied to find a good SPM for the generated STBC and polar code design procedure for the new set-partitioned STBC is repeated. Finally, the best match of the information set, SPM, and parameters of a specific STBC corresponding to the lowest FER is chosen. The design algorithm is formalized in Function Joint-Optim. In this function, $\mathcal{A}$ is the information set, $\mathbf{R}$ is the vector of MLC component code rates, and $\mathbf{z}$ is the vector of multidimensional bit-to-symbol mapping. The modified PSO can be used for the joint optimization by substituting the Function Joint-Optim instead of outage in

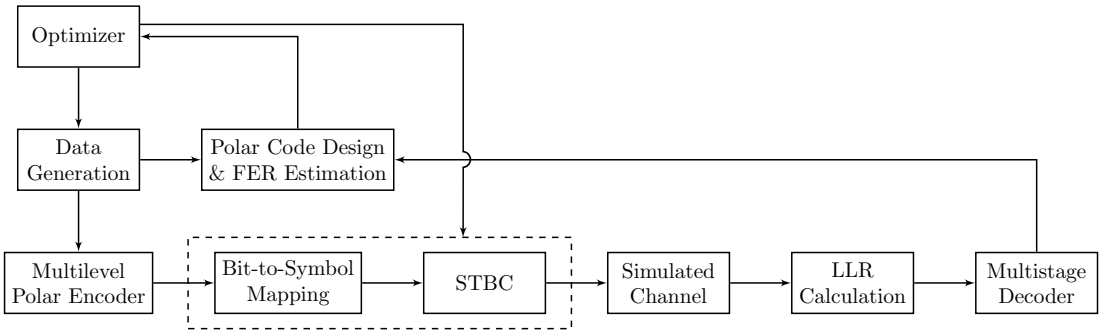



**Figure 7.5:** Joint optimization of polar codes and STBCs block diagram.

Algorithm 8. The block diagram of the joint optimization method is shown in Fig. 7.5.

---

**Function** Joint-Optim($\mathbf{G}_X$,$\boldsymbol{\zeta}$,$R_{tot}$)

---

**Input:** $\mathbf{G}_X$, $R_{tot}$, and $\boldsymbol{\zeta}$
**Output:** $\epsilon$, $\mathcal{A}$ and $\mathbf{R}$
**Procedures:**
Update_Labelling(): Generates the SPM for the input $\mathbf{G}_X$ using the algorithm in Section 3.1.3.
Polarcode_Design(): Designs a multilevel polar code with the total rate $R_{tot}$ for a given constellation and a multidimensional bit-to-symbol mapping using the algorithm in Section 4.2.2 and outputs $\epsilon$, $\mathcal{A}$ and $\mathbf{R}$.

1 $\mathbf{G}_X$ =Update_STBC($\boldsymbol{\zeta}$) ▷ Defined in Algorithm 8.
2 $\mathbf{z}$ =Update_Labelling($\mathbf{G}_X$)
3 $[\epsilon, \mathcal{A}, \mathbf{R}]$ =Polarcode_Design($\mathbf{z}, \mathbf{G}_X, R_{tot}$)
4 return $[\epsilon, \mathcal{A}, \mathbf{R}]$

---

To compare the performance of different design method in Section 7.7, we optimized a few codes using the Function Joint-Optim at a minimum SNR corresponding to a FER of 0.01. The parameters of theses codes are as follows; for $\mathbf{G}_7$, $\alpha_1 = \beta_1 = 0.765$, $\alpha_2 = -0.265 + 0.587\jmath$, and $\beta_2 = 0.587 + 0.265\jmath$; for TV $\mathbf{G}_9$, $\alpha_1 = \alpha_2 = 0.5$ and $\beta_2 = \jmath\beta_1 = -0.410 + 0.287\jmath$; for $\mathbf{G}_{11}$, $\beta_1 = 0.358 + 0.358\jmath$, $\beta_2 = -0.358 - 0.358\jmath$ and $\beta_3 = -0.506$; and for TV $\mathbf{G}_{11}$, $\beta_1 = 0.359 + 0.359\jmath$, $\beta_2 = -0.359 - 0.359\jmath$, and $\beta_3 = -0.507$.

Note that the joint optimization method is of much lower complexity than the

outage probability optimization for short to moderate length codes since the relatively precise numerical evaluation of the outage probability in (7.14) is expensive. However, for long codes, the joint optimization method is expensive and it turns out that the outage probability optimization is cheaper. Furthermore, in real-time optimization when the statistics of the communication channel is not known, both outage probability based optimization and the joint optimization methods can be used. But employing the joint optimization is more affordable since it does not need an additional software as the polar encoder and decoder are embedded in the system.

## 7.7 Numerical Results and Discussions

In this section, the performance of the outage optimized and the joint optimized MLPCM schemes with STBCs and SBCs are compared with MLPCM designed using the rank and determinant criteria, and the mutual information. STBCs are designed based on the rank and determinant criteria, referred to as Method 1, mutual information, referred to as Method 2, the outage probability, referred to as Method 3, and the joint design of FEC codes and STBCs referred to as Method 4. The parameters of the optimized codes are mentioned in Sections 7.3 and 7.6 for Methods 3 and 4, respectively. For designing MLPCM, the construction method in Section 4.5 is employed. We also compare MLPCM with bit-interleaved turbo coded-modulation (BITCM) and bit-interleaved convolutional coded-modulation (BICCM) schemes constructed using a method explained in [81]. For the BITCM, BCJR decoder is used with 20 iterations and for the BICCM the Viterbi decoder is employed. The turbo code in BITCM is LTE's turbo code [14] and the convolutional code generator polynomials are [133 171] with $R_{tot}$ = 0.5 which benefits from a large free distance [81]. The optimized parameters are mentioned in Section 7.3. The polar decoder for all cases is SCD. The SC list decoder, introduced in [62], is also tried but no curve has been

shown. For all SCD curves, SCLD slightly improves the performance, e.g., around 0.2 dB. Unless otherwise stated, we use a $2 \times 2$ antenna configuration. The constellation used for all Matrices $\mathbf{G}_6$, $\mathbf{G}_7$, $\mathbf{G}_9$, and $\mathbf{G}_{11}$ is QPSK and for Matrix $\mathbf{G}_8$ is 16-QAM.

The performance of the BITCM, BICCM, and MLPCM with Matrices A, B, and the Alamouti STBC for 2 bpcu with $R_{tot} = 1/2$ and $N_{tot} = 512$ bits are compared in Fig. 7.6. MLPCM with Alamouti STBC scheme is constructed using an MLC with 4 levels for a 16-QAM constellation with the total code rate of 1/2. Among all curves, the BICCM with $\mathbf{G}_6$ shows the worst performance. The BITCM with $\mathbf{G}_6$ is around 1 dB better than the BICCM but it is 0.8 dB worse than MLPCM with $\mathbf{G}_6$ at a FER of 0.01. Note that the BCJR with 20 iterations is far more complex than the SCD. In the set of curves provided for MLPCM with $\mathbf{G}_7$, the STBC designed using Method 2 is 1.2 dB worse than Method 1 since the mutual information is not an appropriate metric for the slow fading channel. Furthermore, FERs of STBCs designed using Methods 3 and 4 are 0.1 dB and 0.4 dB lower than the one with Method 1, respectively. Thus, as expected, the joint optimization can improve the performance even more than the optimization based on the outage probability.

In Fig. 7.6 for all curves, we used the SPM designed based on the determinant criteria since other metrics do not generate better SPMs or the improvement is negligible. In Fig. 7.7, we evaluate the effect of labelling rules including the Frobenius norm, referred to as Rule 1, modified Frobenius norm in Algorithm 10, referred to as Rule 2, and the design based on the UBPOP, referred to as Rule 3, for 4 bpcu. For all curves, $N_{tot} = 1024$ bits and $R_{tot} = 1/2$. We observe that for MLPCM with $\mathbf{G}_8$, the SPM designed using Rule 2 improves the FER by about 2 dB in comparison to Rule 1 at a FER of 0.01. Moreover for TV $\mathbf{G}_8$, SPMs designed using Rules 2 and 3 work 1.2 and 1.7 dB better than the SPM generated using Rule 1, respectively.

In Fig. 7.8, we provided the compression of $\mathbf{G}_9$ and $\mathbf{G}_8$ at 7.2 bpcu for the total

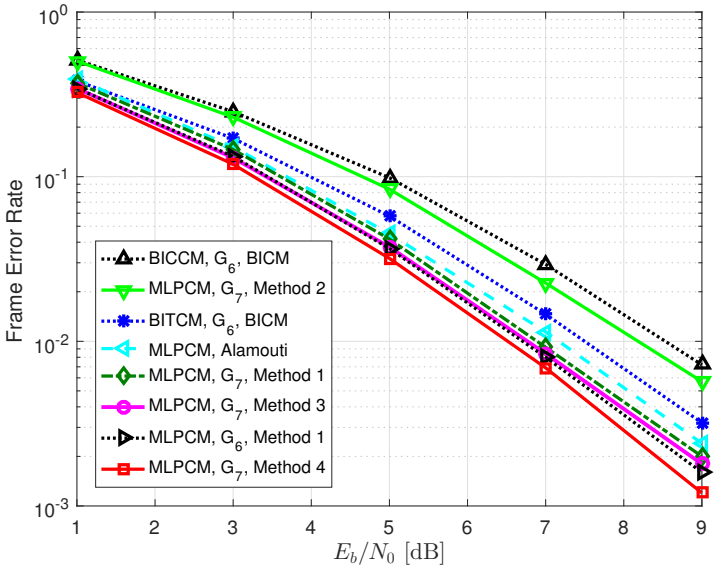



**Figure 7.6:** FER comparison of BICCM, BITCM and MLPCM with $\mathbf{G}_6$, and MLPCM with $\mathbf{G}_7$ designed using Methods 1, 2, 3, and 4, and the Alamouti code for 2 bpcu and $N_{tot} = 512$.

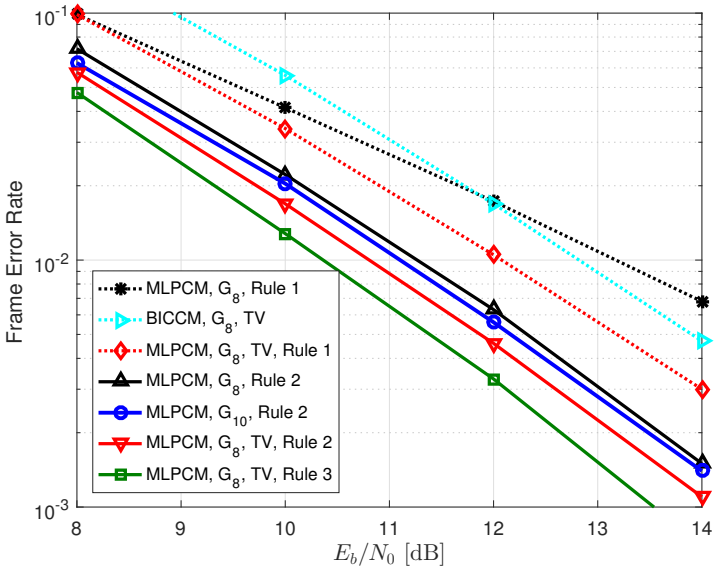



**Figure 7.7:** FER comparison of BICCM and MLPCM with $\mathbf{G}_8$ designed using different rules for SPM generation and $\mathbf{G}_{10}$ designed using Method 3, all for 4 bpcu and $N_{tot} = 1024$.

code length of 128 bits and $R_{tot} = 9/10$. The results indicate that at a FER of 0.001, MLPCM with TV $\mathbf{G}_{10}$ designed using Method 3 outperforms MLPCM with $\mathbf{G}_8$ and TV $\mathbf{G}_8$ by 0.8 dB and 0.3 dB, respectively. Furthermore, for MLPCM with TV $\mathbf{G}_{10}$, optimization using Method 4 provides 0.2 dB improvement over Method 3. It is clear that the joint optimization of polar codes and STBCs for all code rates can slightly improve the performance in comparison to the outage probability based optimization. Also, by comparing Fig 7.8 with Fig. 7.2(b), we realize that the order of curves in terms of the FER is the same as that of the outage probability curves.

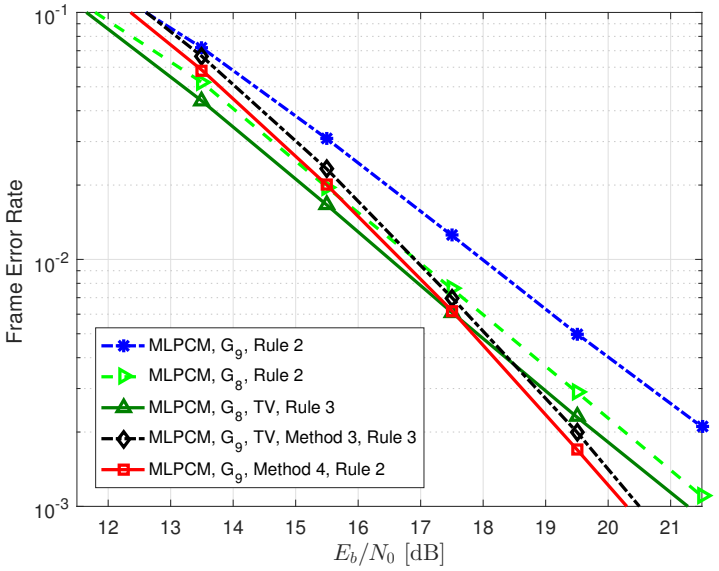



**Figure 7.8:** FER comparison of MLPCM with $\mathbf{G}_8$ and $\mathbf{G}_9$ designed using Methods 3 and 4, all for 7.2 bpcu and $N_{tot} = 128$.

The comparison of MLPCM with $\mathbf{G}_8$ and $\mathbf{G}_{11}$ for 6 bpcu by setting $N_{tot} = 256$, $R_{tot} = 1/2$ and $N_t = N_r = 3$ is shown in Fig. 7.9. We observe that for MLPCM with $\mathbf{G}_8$, employing Rule 2 improves the performance by 0.3 dB in comparison to Rule 1 at a FER of 0.01. This improvement is less than what we have observed for $N_r = 2$ antennas in Fig. 7.7 since as $N_r$ increases, the diversity order increases and eventually the Frobenius norm, as the metric for designing SPM, becomes optimal. In addition,

MLPCM with $\mathbf{G}_{11}$ and TV $\mathbf{G}_{11}$ optimized using Method 4 work 0.5 and 0.6 dB better than MLPCM with $\mathbf{G}_8$ and TV $\mathbf{G}_8$, respectively. Note that we also tried the error bounds in Chapter 6 to optimize $\mathbf{G}_{11}$ but the performance is much worse than the sketched curves since the structure of regular constellations limits the optimization.

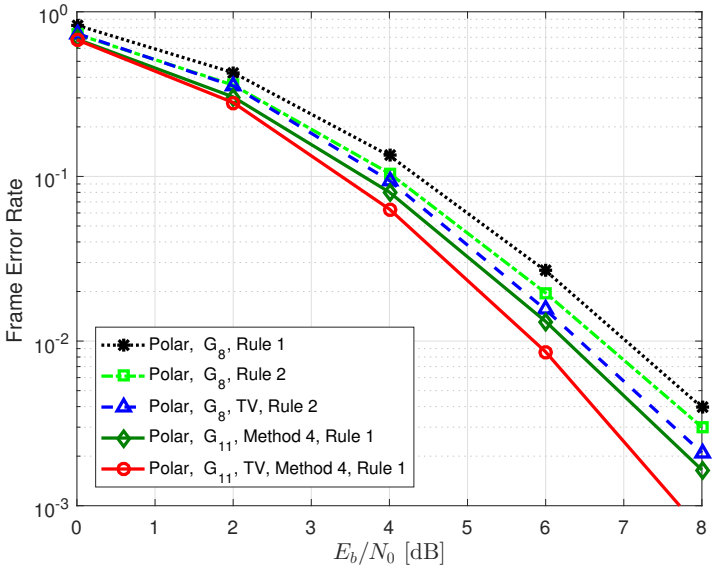



**Figure 7.9:** FER comparison of MLPCM with $\mathbf{G}_8$ and $\mathbf{G}_{10}$ designed using Method 4, all for 6 bpcu and $N_{tot} = 256$ in a $3 \times 3$ antenna configuration.

Finally, for $\mathbf{G}_{12}$ in Fig. 7.10, BIPCM, and MLPCM with independent bit-to-symbol mappings for antenna indices and symbols are compared with MLPCM with the joint bit-to-symbol mapping designed using Rule 1 and Rule 2 for 3 bpcu by setting $R_{tot} = 1/2$, $N_t = 4$, and $N_r = 2$. The constellation used for all antennas is 16-QAM and $N_{tot}$ is set to 1536 and 2048 bits for MLPCM and BIPCM, respectively. We observe that at a FER of 0.01, MLPCM with SPM designed using Rule 1 outperforms MLPCM and BIPCM with independent mappings by 0.4 dB and 0.8 dB, respectively. Also, SPM designed using Rule 2 works better than SPM designed using Rule 1 for 0.9 dB.

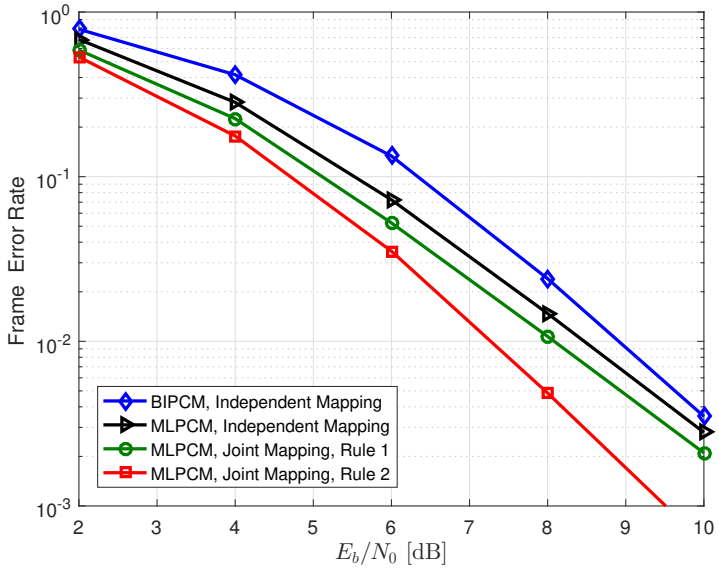



**Figure 7.10:** FER comparison of MLPCM and BICPM with $\mathbf{G}_{12}$ and a variety of SPMs, all for 3 bpcu in a $4 \times 2$ antenna configuration.

## 7.8 Conclusion

In this chapter, we improved the space-time signal design for multilevel polar coding as a low-complexity power-efficient scheme for slow broadcast channels. To do so, we first optimized STBCs for MLPCM by minimizing the outage probability at a target outage or SNR. The method includes limiting the number of free parameter of STBCs by using power and shaping conditions and employing a modified PSO algorithm to find other parameters. In addition, we showed that the outage probability of all levels of MLC/MSD should be optimally equal under certain conditions and based on that we proposed an outage rule for determining component code rates. Furthermore, to design SPM we derived an upper bound on the pairwise outage probability that can model both slow- and fast-changing parameters of the system and the channel. We showed employing this bound to generate SPM for TVSBCs can substantially improve the performance up to 1.7 dB compared to the Frobenius norm. Due to the similarity

of SPMs generated using the derived bound to those generated using the Frobenius norm for SBCs, we further proposed an algorithm to modify pairwise metrics that decreased the FER up to 2 dB.

We also proposed a novel approach to jointly optimize multilevel polar codes and STBCs. For the joint optimization of polar code and STBCs, here, we first change the parameters of a STBC to create a new STBC; then we generate a new SPM using a set-partitioning algorithm and repeat the code design procedure for the new STBC until we find an information set and the corresponding parameters of the STBC that minimize the FER jointly. The modified PSO can be employed for the joint optimization as well. The numerical results show an improvement compared to STBCs, SBCs and TVSBCs designed using the rank and determinant criteria.

## 7.9 Appendix A: Derivation of (7.21)

We begin with the derivation of the Bhattacharyya coefficient for the space-time signal model in (2.1). The pairwise Bhattacharyya coefficient, given the channel matrix $\mathbf{H}$, can be written as

$$\rho(\mathbf{G}_i, \mathbf{G}_j|\bar{\gamma}, \mathbf{H}) = \int_{\mathbf{Y}} \sqrt{\mathrm{P}(\mathbf{Y}|\mathbf{G}_i, \mathbf{H})\mathrm{P}(\mathbf{Y}|\mathbf{G}_j, \mathbf{H})}\, d\mathbf{Y}. \tag{7.26}$$

By substituting (2.2) and using [141, Lemma. 1], (7.26) can be simplified as

$$\exp\Big(\frac{-\mathrm{tr}[\mathbf{H}^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)\mathbf{H}]}{4N_0}\Big). \tag{7.27}$$

Using (7.26), the upper bound on the outage probability denoted as $\mathrm{P}(q_{i,j} < \hat{\rho}_{i,j})$

can be written as

$$\mathrm{P}\left(q_{i,j} < \exp\left(\frac{-\mathrm{tr}[\mathbf{H}^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)\mathbf{H}]}{4N_0}\right)\right). \tag{7.28}$$

Taking ln of $q_{i,j}$ and $\hat{\rho}_{i,j}$, (7.28) can be simplified as

$$\mathrm{P}\left(\mathrm{tr}[\mathbf{H}^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)\mathbf{H}] < -4N_0 \ln(q_{i,j})\right). \tag{7.29}$$

In (7.29), the left hand side of the inequality can be written as

$$\mathrm{tr}[\mathbf{H}^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)^{\mathcal{H}}(\mathbf{G}_i - \mathbf{G}_j)\mathbf{H}] = \sum_{n_r}\sum_{t}\left|\sum_{n_t}\delta_{t,n_t}h_{n_t,n_r}\right|^2. \tag{7.30}$$

Each term $\xi_{t,n_t,n_r} = |\sum_{n_t}\delta_{t,n_t}h_{n_t,n_r}|^2$ in (7.30) can be algebraically extended to the real and imaginary parts. Each real and imaginary part is the weighted sum of terms including real or imaginary part of an element of $\mathbf{H}$ denoted as $h_{n_t,n_r}$. Therefore, both real and imaginary parts are the sum of Gaussian random variables and consequently have Gaussian distribution. Since the real and imaginary parts of $\xi_{t,n_t,n_r}$ are Gaussian and their correlation turns out to be zero, their sum is distributed according to a central chi-squared distribution with two degrees of freedom. In case of SBCs, (7.30) is simplified to $\sum_{n_r}|\sum_{n_t}\delta_{1,n_t}h_{n_t,n_r}|^2$ in which for different receive antennas, the term $\xi_{1,n_t,N_r}$ is independent of others. Therefore, the distribution of sum of $\xi_{1,n_t,N_r}$ is the sum of chi-squared distributions with two degrees of freedom which in turn is a chi-squared distribution with $2N_r$ degrees of freedom and $\mu_2 = \sum_{N_r}\mu_{2,N_r}$ where $\mu_{2,N_r}$ is the variance for each receive antenna. The variance of the each real and imaginary parts of $\xi_{1,n_t,N_r}$ is $\mu_2 = 0.5\sum_{n_t}|\delta_{1,n_t}|^2$. From (7.29), the outage is the CDF of this distribution expressed in (7.21).

## 7.10 Appendix B: Derivation of the Approximation (7.22)

In this section, we use the definitions of Appendix. 7.9. In case of STBCs, for each receive antenna, (7.30) can be simplified as $\sum_{t} \xi_{t,n_t,n_r}$ which includes $T$ correlated terms. Since $\xi_{t,n_t,n_r}$ has gamma distribution, for $T = 2$, $\sum_{t} \xi_{t,n_t,n_r}$ is distributed according to a type I McKay distribution [174]. For $T > 2$, the general expression $\sum_{t} \xi_{t,n_t,n_r}$ is the sum of correlated gamma random variables and is distributed according to a complex infinite power series [175]. As a computationally cheap yet accurate approximation, the sum of correlated gamma random variables can be accurately approximated as a gamma random variable by matching the first two moments [176]. In a few straightforward steps, these moments can be derived as 7.23 and the CDF is expressed in (7.22).

## 7.11 Appendix C: Derivation of the Approximation (7.24)

The Bhattacharyya coefficient in (7.27) for one receive antenna by considering TV sequences can be expressed as

$$\rho(\mathbf{G}_i, \mathbf{G}_j|\bar{\gamma}, \mathbf{H}, \boldsymbol{\theta}) = \exp\Big( - \frac{1}{4\mathrm{N}_0} \Big| \sum_{\mathrm{n_t}} \delta_{1,\mathrm{n_t}} \mathrm{h}_{\mathrm{n_t},\mathrm{n_r}} \mathrm{e}^{\jmath\theta_{\mathrm{n_t}}} \Big|^2 \Big). \tag{7.31}$$

Due to the fast-changing nature of TV sequences and assuming the code length tends to infinity, the integral over $\rho(\mathbf{G}_i, \mathbf{G}_j|\bar{\gamma}, \mathbf{H}, \boldsymbol{\theta})$ can be taken as

$$\begin{aligned}&\int_{\boldsymbol{\theta}} \exp\Big(-\frac{1}{4N_0}\Big|\sum_{n_t}\delta_{t,n_t}h_{n_t,n_r}e^{j\theta_{n_t}}\Big|^2\Big)\frac{d\boldsymbol{\theta}}{(2\pi)^{N_t}}\\ &= \exp\Big(-\frac{1}{4N_0}\sum_{n_t}|\delta_{1,n_t}h_{n_t,n_r}|^2\Big)\int_{\boldsymbol{\theta}}\frac{\exp(-F_{n_r}(\Delta_{i,j}))}{(2\pi)^{N_t}}d\boldsymbol{\theta},\end{aligned} \tag{7.32}$$

where

$$F_{n_r}(\Delta_{i,j}) = \frac{2}{4N_0}Re\{\sum_{u=1}^{N_t}\sum_{v=1}^{N_t}\delta_{1,u}\delta_{1,v}h_{u,n_r}h_{v,n_r}e^{-j\hat{\theta}}\}, \tag{7.33}$$

in which $\hat{\theta} = \theta_u - \theta_v$ and $(2\pi)^{-N_t}$ is the density function of $\boldsymbol{\theta}$ which is uniformly distributed in range $[0, 2\pi]$. Taking the general form of the integral $\int_{\boldsymbol{\theta}}\exp(-\frac{1}{4N_0}F_{n_r}(\Delta_{i,j}))$ in (7.32) is difficult. However, for $N_t = 2$, it can be simplified as

$$I_0(\frac{1}{2N_0}|\delta_{1,1}\delta_{1,2}|), \tag{7.34}$$

where $I_0$ is the modified Bessel function of the first kind with zero order. To estimate the upper bound on the outage probability, we take the logarithm of the Bhattacharyya coefficient in 7.29. Similarly, for $N_t = 2$ and $N_r = 1$ by taking the logarithm, the left hand side of 7.29 is simplified as

$$\frac{1}{4N_0}\underbrace{\sum_{n_t}|\delta_{1,n_t}h_{n_t,n_r}|^2}_{\text{Part 1}} - \underbrace{\ln\Big(I_0(\frac{1}{2N_0}|\prod_{n_t}\delta_{1,n_t}h_{n_t,n_r}|)\Big)}_{\text{Part 2}}. \tag{7.35}$$

In (7.35), $\ln\big(I_0(x)\big)$ can be estimated using piecewise linear approximation given as

$$\ln\big(I_0(x)\big) \approx a_1x + a_2 = \begin{cases}0.12x & 0 < x \le 0.5,\\ 0.35x - 0.12 & 0.5 < x \le 1,\\ 0.59x - 0.37 & 1 < x \le 2,\\ 0.8x - 0.81 & 2 < x \le 4,\\ 0.92x - 1.3 & x \ge 4.\end{cases} \tag{7.36}$$

Note that for $x > 4$, the linear approximation remains valid for a wide range and thus, for designing a SPM, $\delta_{1,n_t}$ or the signal energy can be adjusted to lie in this region.

In (7.35), Part 1 is the sum of independent chi-squared random variables since $\delta_{1,1}h_{1,n_t}$ can be separated into real and imaginary parts each including a Gaussian random variable. The sum of independent chi-squared random variables is chi-squared which is a special case of the gamma distribution. Thus, Part 1 is a gamma random variable with $\mu_1 = \sum_{n_t} |\delta_{1,n_t}|^2$ and $\mu_2 = 2\sum_{n_t} |\delta_{1,n_t}|^4$. Using the linear approximation in (7.36), Part 2 in (7.35) is an affine transform of $\Upsilon = |\prod_{n_t} \delta_{1,n_t} h_{1,n_t}|$. $\Upsilon$ is the product of two Rayleigh random variables and is distributed according to a double Rayleigh [177, Eq. (11)]. The double Rayleigh distribution can in turn be represented as $f(x) = \frac{\pi^4 x}{(4p_1)^2} K_0(\frac{x}{p_1})$, where the parameter $p_1$ equals $\prod_{n_t} E[|h_{1,nt}|]|\delta_{1,n_t}|$ and $K_0(.)$ is the modified Bessel function of the second kind with zero order. Thus, (7.35) is the difference of two correlated gamma and double Rayleigh random variables. This distribution is unknown but we can approximate it by a well-known distribution. To find a close distribution for a given pair of STBC points, the empirical distribution in (7.35) is generated and compared with a large number of classes of distributions and the one with the lowest Kullback-Leibler divergence is chosen. To find a distribution that works well on average for all points of the STBC, the average Kullback-Leibler divergence over all STBC points is used. Thus, the metric to choose the distribution is given as

$$\frac{1}{2^{2B}} \sum_u \sum_v \sum_i \mathrm{P}_i(\mathbf{G}_u, \mathbf{G}_v) \log_2 \left( \frac{\mathrm{P}_i(\mathbf{G}_u, \mathbf{G}_v)}{\mathrm{Q}_i(\mathbf{G}_u, \mathbf{G}_v)} \right), \tag{7.37}$$

where $\mathrm{P}_i(\mathbf{G}_u, \mathbf{G}_v)$ and $\mathrm{Q}_i(\mathbf{G}_u, \mathbf{G}_v)$ are the $i^{th}$ elements of the vector of probability values of the tested and empirical distributions, respectively. After testing a large number of classes of one and two parameters distributions, the log-normal distribution

is chosen to model the distribution in (7.35). To find the parameters of the under-test distribution, we match the first two moments. Note that although three parameter distributions may better model the empirical data, matching their first two moments may be difficult. For $N_r > 1$, Part 1 in (7.35) changes to $\frac{1}{4N_0}\sum_{n_r}\sum_{n_t}|\delta_{1,n_t}h_{n_t,n_r}|^2$ and Part 2 should be extended as $-\ln(I_0(\frac{1}{2N_0}|\sum_{n_r}\delta_{1,1}h_{1,n_r}\delta^*_{1,2}h^*_{2,n_r}|))$. The matched CDF and the first two moments are presented in (7.24) and (7.25), respectively. The empirical and the matched log-normal CDFs are depicted in Fig. 7.11 for two SNRs.

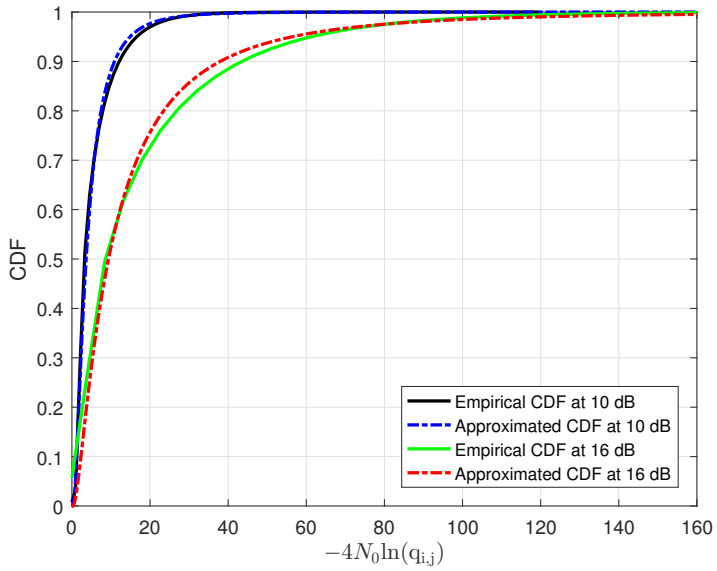



**Figure 7.11:** Comparison of the empirical CDF and the approximated CDF for a given pair of TVSBC points at two different SNRs and $N_r = 1$.

# Chapter 8

# Conclusions and Future Work

In this thesis, MLPCM is designed and analyzed as a low-complexity high-performance coded-modulation scheme for wireless channels. In particular, in Chapter 3, the LLR complexity for QAM constellations is highly simplified by employing a novel device. The proposed device consists of a simple linear block code and two PAMs for modulating the real and imaginary parts of the QAM. In addition, a modified binary switching algorithm for designing Gray-like mapping for BICM is proposed.

In Chapter 4, a low-complexity GA-based approach is explained to design polar coded-modulation for the AWGN channel by minimizing the FER at a given SNR. This method uses the simple LLR estimation for QAM proposed in Chapter 3 to estimate the average LLR used in code design procedure. Also, a GA-based method is proposed to design polar coded-modulation for orthogonal transmission in slow fading channels. This method consists of designing polar codes using GA for a specific $\mathbf{H}$ realization and repeating the design until the information set does not change anymore. In practice, we observe that with only a small number of $\mathbf{H}$ realizations (e.g., 35 realizations), polar codes similar to those designed using the simulation-based approach can be constructed. In addition, a low-memory-space/low-arithmetic-complexity polar code design method based on modeling the bit-channel error at

different SNRs in an AWGN channel is proposed that can be used for the real-time constructing and fast adapting of polar codes. The same method is extended to the case of the slow fading channel and is simplified. Numerical results showed that the performance of codes designed using the proposed method is the same as codes designed using the simulation-based approach.

In Chapter 5, the polar code and MLPCM design based on the throughput for SCD were proposed. Throughput-based design of polar codes can be employed in code design for AMC and HARQ techniques. To design this class of polar codes, a set of algorithms is presented and codes were designed for NC and CC HARQ protocols. Furthermore, polar codes are designed by maximizing the throughput for the SCLD. The results showed that polar codes designed using this method can work close to the channel capacity.

In Chapter 6, multidimensional constellations are compared with STBCs in terms of the FER for slow fading channels. To this end, different methods of the optimization of uncoded high-SNR multidimensional constellations, based on the two novel upper bounds on the SER and BER in Nakagami-$m$ fading channel, are discussed and it is shown that the optimized constellations improve the performance in comparison to the best-known constellations in literature. Due to the ability of these constellations to be designed for a wide range of spectral efficiencies, they can improve the performance of coded-modulation schemes in the poor performance region of available constellations and schemes, e.g., at low spectral efficiencies. However, at high spectral efficiencies, outage optimized STBCs or STBCs designed using the rank and determinant criteria show better performance.

Finally, in Chapter 7, the MLPCM scheme is designed for slow broadcast channels. To this end, due to achieving the channel capacity by MSD, a method for optimizing STBCs at low-to-moderate SNRs is proposed using the PSO. This method slightly improves the FER of MLPCM in comparison to the rank and determinant criteria for

STBCs. However, for cases that the rank and determinant criteria or other high-SNR criteria cannot be employed, e.g., for many SBCs and TVSBCs, the outage-based optimization is the only available design method. To design SPM for slow fading channels, a bound on the outage probability is derived by substituting the cutoff rate instead of the mutual information in outage probability definition. This bound specifically can model both fast- and slow-changing features of the channel. Using this bound, approximate closed-form expressions for the pairwise outage probability of SBCs and TVSBCs is achieved that in case of TVSBCs can improve the FER of MLPCM by 1.7 dB. The proposed bound does not improve the SPM compared to the Frobenius norm and the determinant criterion for SBCs and STBCs, respectively. Thus, an algorithm is proposed for modifying pairwise metrics for SBCs that can improve the performance of SBCs up to 2 dB. In addition, it was shown that the outage probability of all levels of MLC/MSD are optimally equal if for all realizations of $\mathbf{H}$ either all levels are in outage or no level is in outage. By assuming the equal outage probability for all levels given that we achieve the target spectral efficiency, the component code rates of MLPCM are determined. Numerical results showed that this rule works the same as the best bit-channel rule.

The joint optimization of polar codes and STBCs based on the FER, as a novel method of designing STBCs concatenated with MLPCM, is also proposed in Chapter 7. Indeed, mutual information or the outage probability as traditional methods of designing high-performance constellations for FEC coded systems cannot adapt the constellation to the structure of the FEC code or the bit-to-symbol mapping. Therefore, the constellation or STBC should be optimized based on bounds on the performance of the joint system, which are difficult to derive for most modern FEC codes, e.g., LDPC codes. However, due to achieving a low-complexity estimation of the FER during the simulation-based design of polar coded-modulation, we proposed a method for the joint optimization of polar codes and STBCs that improves the

performance of the joint system. The proposed method includes generating a set of parameters for a specific STBC, generating a bit-to-symbol mapping, designing polar code for the generated signal and repeating these steps until the best match of the STBC and the polar code is achieved. Note that the simulation-based design of MLPCM plays the role of a bound since this approach can approximate the average FER of the scheme for any given channel and modulation.

## 8.1 Future Work

The works and results of this thesis raise important and interesting problems for research. Here, we review selected problems briefly.

**Improving metrics for designing multidimensional constellations:** Error bounds were used in Chapter 6 for designing multidimensional constellations. However, their performance is not as good as the STBC design based on the outage probability when used with FEC codes. Multidimensional constellations benefit from a larger number of free parameters compared to STBCs and their optimization may result in higher performance signaling compared to STBCs. Thus, finding new computationally cheap metrics for designing multidimensional constellations is of importance. To this end, features of the outage optimized constellations at low-to-moderate SNRs should be studied and based on them, novel metrics and constellation structures may be proposed to improve the signal design.

**Designing MLPCM for a large number of transmit antennas:** Throughout this thesis, we only presented the MLPCM design for the AWGN channel and MIMO schemes in slow fading channels. These cases can be employed in a wide range of applications, e.g., the machine-to-machine communication, the Internet of Things, the control channel, and the data channel in wireless networks. Recently, massive MIMO with millimeter wave technology is proposed for the line-of-sight data channel in 5G

and beyond. Due to the low-complexity, the high-performance, and the adaptability of MLPCM, it can be designed and employed for massive MIMO systems in 5G and beyond.

**Optimizing the limited feedback codebook for slow fading channels:** In Chapter 7, we optimized the signal for transmission in slow fading channels by minimizing the outage probability when CSIT is not available. When partial CSIT is available, the signal can be optimized based on outage by considering the final throughput. The codebook structure depends on the feedback rate and the SNR. This class of codebook design can highly improve the performance of the limited feedback MLPCM schemes.

**Designing throughput-optimal polar codes for Type II HARQ:** The throughput-based polar code design is proposed in Chapter 5 for AMC, NC and CC HARQ protocols. In Type-II HARQ systems, to improve the throughput, every time a repeat request is received only a new set of parity bits, instead of the whole codeword, is transmitted which is referred to as IR. For Type-II HARQ, typically incremental redundancy codes are employed, e.g., the design of incremental redundancy turbo and LDPC codes are reported in [178–180] and the incremental redundancy polar codes are proposed in [116,181]. One of the shortcomings of currently available Type-II HARQ schemes is the difficulty in scheduling. Due to the substantial throughput improvement that can be achieved using carefully designed codes, the idea of throughput-based polar code design should be extended to designing optimal codes for IR HARQ systems in the future.

**Throughput-based joint optimization of polar codes and MIMO schemes:** Short-to-moderate length codes can be employed with limited feedback MIMO schemes for practical application. In this case, polar codes and MIMO schemes can be jointly optimized using the method proposed in Chapter 7.

**Partial CSIR MLPCM design in slow fading channels:** For all designs in

this thesis, availability of full CSIR is considered. However, in many practical scenarios, even in slow fading channels, only partial CSIR is available. Thus, extending the bounds on outage probability derived in this thesis to design SPM is of importance. Furthermore, finding low memory-space/low-arithmetic-complexity methods for designing polar codes under imperfect CSIR is an open problem.

# List of References

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, pp. 3051–3073, July 2009.

[2] E. Şaşoğlu and L. Wang, "Universal polarization," *IEEE Transactions on Information Theory*, vol. 62, pp. 2937–2946, June 2016.

[3] A. Alamdar-Yazdi and F. R. Kschischang, "A simplified successive-cancellation decoder for polar codes," *IEEE Communications Letters*, vol. 15, pp. 1378–1380, December 2011.

[4] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check," *IEEE Communications Letters*, vol. 16, pp. 2044–2047, December 2012.

[5] C. Leroux, A. J. Raymond, G. Sarkis, and W. J. Gross, "A semi-parallel successive-cancellation decoder for polar codes," *IEEE Transactions on Signal Processing*, vol. 61, pp. 289–299, January 2013.

[6] G. Sarkis, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Fast polar decoders: Algorithm and implementation," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 946–957, May 2014.

[7] G. Sarkis, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Fast list decoders for polar codes," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 318–328, February 2016.

[8] J. L. Massey, "Coding and modulation in digital communications," in *Proc. of International Zurich Seminar on Digital Communications*, (Zurich, Switzerland), March 1974.

[9] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, pp. 927–946, May 1998.

[10] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, pp. 371–377, May 1977.

[11] H. Imai and S. Hirakawa, "Correction to 'a new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, p. 784, November 1977.

[12] S. Y. Le Goff, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Transactions on Wireless Communications*, vol. 6, June 2007.

[13] ETSI, *Universal mobile telecommunications system (UMTS); multiplexing and channel coding (FDD), 3GPP TS 25.212, version 8.5.0*, March 2009.

[14] ETSI, *LTE; Evolved universal terrestrial radio access (E-UTRA); Multiplexing and channel coding, 3GPP TS 36.212, version 12.8.0*, March 2016.

[15] IEEE, *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5GHz band, IEEE 802.11*, July 1999.

[16] IEEE, *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications Amendment 5: Enhancements for higher throughput, IEEE 802.11n*, October 2009.

[17] ETSI, *Digital video broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2), EN 302 755 V1.1.1*, September 2009.

[18] ETSI, *Digital video broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital transmission system for cable system (DVB-C2), EN 302 769 V1.1.1*, April 2010.

[19] ETSI, *Digital video broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2), EN 302 307 V1.2.1*, August 2009.

[20] ETSI, *Digital Radio Mondiale (DRM); System Specification, ES 201 980 V3.1.1*, August 2009.

[21] U. Wachmann, J. B. Huber, and P. Schramm, "Comparison of coded modulation schemes for the AWGN and the Rayleigh fading channel," in *Proc. of IEEE International Symposium on Information Theory*, (Cambridge, MA), August 1998.

[22] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 2141–2155, September 2003.

[23] A. Elkhazin, *Space-time Coded Modulation Design in Slow Fading*. PhD thesis, University of Toronto, 2009.

[24] K. Chen, K. Niu, and J.-R. Lin, "An efficient design of bit-interleaved polar coded modulation," in *Proc. of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications*, (London, UK), September 2013.

[25] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Multilevel polar-coded modulation," in *Proc. of IEEE International Symposium on Information Theory*, (Munich, Germany), January 2013.

[26] C. I. Ionita, M. Mansour, J. C. Roh, and S. Hosur, "On the design of binary polar codes for high-order modulation," in *Proc. of IEEE Global Communications Conference*, (Austin, TX, USA), December 2014.

[27] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, June 1996.

[28] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, November 1999.

[29] J. E. Uddenfeldt and A. K. Raith, *Cellular digital mobile radio system and method of transmitting information in a digital cellular mobile radio system*. US patent no. 5,088,108, February 1992.

[30] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. of IEEE International Conference on Communications*, (Geneva, Switzerland), May 1993.

[31] G. Golden, C. Foschini, R. A. Valenzuela, and P. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, pp. 14–16, January 1999.

[32] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1841–1852, November 1999.

[33] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[34] ETSI, *LTE; Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation, 3GPP TS 36.211, version 12.8.0*, December 2015.

[35] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A $2 \times 2$ full-rate space-time code with nonvanishing determinants," *IEEE Transactions on Information Theory*, vol. 51, pp. 1432–1436, April 2005.

[36] F. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Transactions on Information Theory*, vol. 52, p. 3885, September 2006.

[37] D. Duyck, S. Yang, F. Takawira, J. J. Boutros, and M. Moeneclaey, "Time-varying space-only code: A new paradigm for coded MIMO communication," in *Proc. of International Symposium on Communications Control and Signal Processing*, (Rome, Italy), May 2012.

[38] D. Duyck, M. Moeneclaey, S. Yang, F. Takawira, and J. J. Boutros, "Time-varying space-only codes," in *Proc. of IEEE International Symposium on Information Theory*, (Cambridge, MA, USA), July 2012.

[39] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), November 2001.

[40] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1051–1064, May 2002.

[41] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2735–2747, October 2003.

[42] T. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, pp. 1458–1461, October 1999.

[43] R. W. Heath, S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Communications Letters*, vol. 5, pp. 142–144, April 2001.

[44] R. Heath and A. Paulraj, "Antenna selection for spatial multiplexing systems based on minimum error rate," in *Proc. of IEEE International Conference on Communications*, (Helsinki, Finland), June 2001.

[45] D. A. Gore and A. J. Paulraj, "MIMO antenna subset selection with space-time coding," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2580–2588, October 2002.

[46] S. M. Alamouti, V. Tarokh, and P. Poon, "Trellis-coded modulation and transmit diversity: Design criteria and performance evaluation," in *Proc. of IEEE International Conference on Universal Personal Communications*, (Florence, Italy), October 1998.

[47] G. Bauch, "Concatenation of space-time block codes and "turbo"-TCM," in *Proc. of IEEE International Conference on Communications*, (Vancouver, Canada), June 1999.

[48] S. Haykin, M. Sellathurai, Y. de Jong, and T. Willink, "Turbo-MIMO for wireless communications," *IEEE Communications Magazine*, vol. 42, pp. 48–53, October 2004.

[49] Ü. Özgür, *A Performance Comparison of Polar Codes with Convolutional Turbo Codes*. PhD thesis, Bilkent University, 2009.

[50] S. Zhao, Q. Sun, M.-K. Feng, and B. Zheng, "A concatenation scheme of polar codes and space-time block codes in multiple-input multiple-output channels," in *Proc. of International Congress on Image and Signal Processing*, (Hangzhou, China), December 2013.

[51] P. Akuon and H. Xu, "Polar coded spatial modulation," *IET Communications*, vol. 8, pp. 1459–1466, April 2014.

[52] P. R. Balogun, I. D. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Polar codes for noncoherent MIMO signalling," in *Proc. of IEEE International Conference on Communications*, (Kuala Lumpur, Malaysia), May 2016.

[53] M. F. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *Proc. of IEEE International Symposium on Information Theory*, (Nice, France), June 2007.

[54] R. Krishnan, A. G. i Amat, T. Eriksson, and G. Colavolpe, "Constellation optimization in the presence of strong phase noise," *IEEE Transactions on Communications*, vol. 61, pp. 5056–5066, October 2013.

[55] A. El Falou, C. Langlais, C. A. Nour, and C. Douillard, "Low ML-detection complexity, adaptive $2 \times 2$ STBC, with powerful FEC codes," in *Proc. of International Symposium on Turbo Codes and Iterative Information Processing*, (Gothenburg, Sweden), August 2012.

[56] H. Jafarkhani, *Space-time Coding: Theory and Practice*. Cambridge, UK: Cambridge University Press, 2005.

[57] D. Duyck, J. J. Boutros, and M. Moeneclaey, "Precoding for outage probability minimization on block fading channels," *IEEE Transactions on Information Theory*, vol. 59, pp. 8250–8266, September 2013.

[58] E. Arıkan, "Channel combining and splitting for cutoff rate improvement," *IEEE Transactions on Information Theory*, vol. 52, pp. 628–639, February 2006.

[59] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets," *IEEE Communications Magazine*, vol. 25, pp. 12–21, February 1987.

[60] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Transactions on Information Theory*, vol. 59, pp. 6562–6582, September 2013.

[61] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, pp. 429–445, March 1996.

[62] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, pp. 2213–2226, May 2015.

[63] L.-J. Lampe, R. Schober, and R. F. Fischer, "Multilevel coding for multiple-antenna transmission," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 203–208, January 2004.

[64] P. A. Martin, D. M. Rankin, and D. P. Taylor, "Multi-dimensional space-time multilevel codes," *IEEE Transactions on Wireless Communications*, vol. 5, November 2006.

[65] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Transactions on Information Theory*, vol. 57, pp. 6650–6672, October 2011.

[66] A. Alvarado, L. Szczecinski, E. Agrell, and A. Svensson, "On the design of interleavers for BICM transmission," in *Proc. of European Wireless Conference*, (Prague, Czech Republic), June 2008.

[67] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Aspects of polar-coded modulation," in *Proc. of International ITG Conference on Systems, Communication and Coding*, (Munich, Germany), January 2013.

[68] J. B. Huber and U. Wachsmann, "Capacities of equivalent channels in multilevel coding schemes," *Electronics Letters*, vol. 30, pp. 557–558, March 1994.

[69] J. B. Huber, "Multilevel codes: Distance profiles and channel capacity," *ITG-Fachbericht*, vol. 130, pp. 305–319, October 1994.

[70] A. B. H. Fredj and J.-C. Belfiore, "LLR computation for multistage decoding," in *Proc. of IEEE European Conference on Networks and Communications*, (Athens, Greece), June 2016.

[71] Y. Tong, T. H. Yeap, and J.-Y. Chouinard, "VHDL implementation of a turbo decoder with log-MAP-based iterative decoding," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, pp. 1268–1278, August 2004.

[72] R. Pyndiah, A. Picart, and A. Glavieux, "Performance of block turbo coded 16-QAM and 64-QAM modulations," in *Proc. of Global Communications Conference*, (Singapore), November 1995.

[73] G. Gül, A. Vargas, W. H. Gerstacker, and M. Breiling, "Low complexity demapping algorithms for multilevel codes," *IEEE Transactions on Communications*, vol. 59, pp. 998–1008, April 2011.

[74] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, pp. 389–399, March 2003.

[75] R. Wang and G. B. Giannakis, "Approaching MIMO channel capacity with soft detection based on hard sphere decoding," *IEEE Transactions on Communications*, vol. 54, pp. 587–590, April 2006.

[76] G. D. Forney, "Coset codes. I: Introduction and geometrical classification," *IEEE Transactions on Information Theory*, vol. 34, pp. 1123–1151, September 1988.

[77] K. Chen, K. Niu, and J. Lin, "Polar coded modulation with optimal constellation labeling," in *Proc. of National Doctoral Academic Forum on Information and Communications Technology*, (Beijing, China), August 2013.

[78] A. Alvarado, E. Agrell, L. Szczecinski, and A. Svensson, "Exploiting UEP in QAM-based BICM: Interleaver and code design," *IEEE Transactions on Communications*, vol. 58, February 2010.

[79] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Transactions on Information Theory*, vol. 42, pp. 409–428, March 1996.

[80] H. Afşer, N. Tirpan, H. Deliç, and M. Koca, "Bit-interleaved polar-coded modulation," in *Proc. of IEEE Wireless Communications and Networking Conference*, (Istanbul, Turkey), April 2014.

[81] L. Mroueh, S. Rouquette-Leveil, and J.-C. Belfiore, "Application of perfect space time codes: PEP bounds and some practical insights," *IEEE Transactions on Communications*, vol. 60, pp. 747–755, March 2012.

[82] R. Yazdani and M. Ardakani, "Efficient LLR calculation for non-binary modulations over fading channels," *IEEE Transactions on Communications*, vol. 59, pp. 1236–1241, May 2011.

[83] G. D. Forney and L.-F. Wei, "Multidimensional constellations. I: Introduction, figures of merit, and generalized cross constellations," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 877–892, August 1989.

[84] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Transactions on Communications*, vol. 38, pp. 2147–2158, December 1990.

[85] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Communications Letters*, vol. 7, pp. 593–595, December 2003.

[86] Y. Li and X.-G. Xia, "Constellation mapping for space-time matrix modulation with iterative demodulation/decoding," *IEEE Transactions on Communications*, vol. 53, pp. 764–768, May 2005.

[87] S. H. Hassani and R. Urbanke, "Universal polar codes," in *Proc. of IEEE International Symposium on Information Theory*, (Honolulu, HI, USA), June 2014.

[88] S. Hassani, S. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proc. of Allerton Conference on Communication, Control, and Computing*, (Monticello, IL, USA), September 2009.

[89] V. Bioglio, C. Condo, and I. Land, "Design of polar codes in 5G new radio," *arXiv:1804.04389*, April 2018.

[90] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Communications Letters*, vol. 13, pp. 519–521, July 2009.

[91] D. Kern, S. Vorköper, and V. Kühn, "A new code construction for polar codes using min-sum density," in *Proc. of International Symposium on Turbo Codes and Iterative Information Processing*, (Bremen, Germany), August 2014.

[92] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Transactions on Communications*, vol. 60, pp. 3221–3227, November 2012.

[93] H. Vangala, E. Viterbo, and Y. Hong, "A comparative study of polar code constructions for the AWGN channel," *arXiv:1501.02473*, January 2015.

[94] A. Bravo-Santos, "Polar codes for the Rayleigh fading channel," *IEEE Communications Letters*, vol. 17, pp. 2352–2355, December 2013.

[95] P. Trifonov, "Design of polar codes for Rayleigh fading channel," in *Proc. of International Symposium on Wireless Communication Systems*, (Brussels, Belgium), August 2015.

[96] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEÜ - International Journal of Electronics and Communications*, vol. 54, pp. 389–398, November 2000.

[97] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 944–957, May 2001.

[98] C. Fragouli, R. D. Wesel, D. Sommer, and G. Fettweis, "Turbo codes with non-uniform constellations," in *Proc. of IEEE International Conference on Communications*, (Helsinki, Finland), June 2001.

[99] B. Lu, G. Yue, and X. Wang, "Performance analysis and design optimization of LDPC-coded MIMO OFDM systems," *IEEE Transactions on Signal Processing*, vol. 52, pp. 348–361, February 2004.

[100] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Transactions on Communications*, vol. 52, pp. 670–678, April 2004.

[101] H. Sankar, N. Sindhushayana, and K. R. Narayanan, "Design of low-density parity-check (LDPC) codes for high order constellations," in *Proc. of IEEE Global Telecommunications Conference*, (Dallas, TX, USA), November 2004.

[102] H. Afşer, N. Tirpan, H. Deliç, and M. Koca, "Bit-interleaved polar-coded modulation," in *Proc. of IEEE Wireless Communications and Networking Conference*, (Istanbul, Turkey), April 2014.

[103] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Polar coding for bit-interleaved coded modulation," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 3115–3127, May 2016.

[104] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Polar-coded modulation," *IEEE Transactions on Communications*, vol. 61, pp. 4108–4119, October 2013.

[105] S. Sun and Z. Zhang, "Designing practical polar codes using simulation-based bit selection," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, pp. 594–603, December 2017.

[106] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, pp. 1361–1391, July 1999.

[107] S.-Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, pp. 657–670, February 2001.

[108] S. B. Korada, A. Montanari, E. Telatar, and R. Urbanke, "An empirical scaling law for polar codes," in *Proc. of IEEE International Symposium on Information Theory*, (Austin, TX, USA), June 2010.

[109] M. Mondelli, S. H. Hassani, and R. Urbanke, "Construction of polar codes with sublinear complexity," in *Proc. of IEEE International Symposium on Information Theory*, (Aachen, Germany), June 2017.

[110] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.

[111] ETSI, *LTE; Evolved universal terrestrial radio access (E-UTRA); Medium access control (MAC) protocol specification, 3GPP TS 36.321, version 12.9.0*, March 2016.

[112] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.

[113] D. Divsalar and F. Pollara, "On the design of turbo codes," Tech. Rep. TDA 42-123, Jet Propulsion Laboratory, Pasadena, CA, USA, November 1995.

[114] D. J. MacKay and R. M. Neal, "Near shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 32, pp. 1645–1646, August 1996.

[115] R. H. Deng and M. L. Lin, "A type I hybrid ARQ system with adaptive code rates," *IEEE Transactions on Communications*, vol. 43, pp. 733–737, February/March/April 1995.

[116] H. Saber and I. Marsland, "An incremental redundancy hybrid ARQ scheme via puncturing and extending of polar codes," *IEEE Transactions on Communications*, vol. 63, pp. 3964–3973, November 2015.

[117] H. Liang, A. Liu, Y. Zhang, and Q. Zhang, "Analysis and adaptive design of polar coded HARQ transmission under SC-List decoding," *IEEE Wireless Communications Letters*, vol. 6, pp. 798–801, December 2017.

[118] K. Chen, K. Niu, Z. He, and J. Lin, "Polar coded HARQ scheme with chase combining," in *Proc. of IEEE Wireless Communications and Networking Conference*, (Istanbul, Turkey), April 2014.

[119] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 757–764, March 2005.

[120] M. El-Khamy, H. P. Lin, J. Lee, H. Mahdavifar, and I. Kang, "HARQ rate-compatible polar codes for wireless channels," in *Proc. of IEEE Global Communications Conference*, (San Diego, CA, USA), December 2015.

[121] Q. Luo and P. Sweeney, "Hybrid-ARQ protocols based on multilevel coded modulation," *Electronics Letters*, vol. 39, pp. 1063–1065, July 2003.

[122] R. A. Thisted, *Elements of Statistical Computing: Numerical Computation*, vol. 1. USA: Chapman and Hall/CRC, 1988.

[123] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, pp. 357–366, March 1998.

[124] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[125] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, vol. 290. Springer Science & Business Media, 2013.

[126] N. J. A. Sloane, R. H. Hardin, T. D. S. Duff, and J. H. Conway, "Minimal-energy clusters of hard spheres," *Discrete & Computational Geometry*, vol. 14, pp. 237–259, October 1995.

[127] M. Beko and R. Dinis, "Designing good multi-dimensional constellations," *IEEE Wireless Communications Letters*, vol. 1, pp. 221–224, June 2012.

[128] A. B. Sediq, P. Djukic, H. Yanikomeroglu, and J. Zhang, "Optimized nonuniform constellation rearrangement for cooperative relaying," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 2340–2347, June 2011.

[129] A. Yadav, M. Juntti, and J. Lilleberg, "Partially coherent constellation design and bit-mapping with coding for correlated fading channels," *IEEE Transactions on Communications*, vol. 61, pp. 4243–4255, October 2013.

[130] A. Gersho and V. Lawrence, "Multidimensional signal constellations for voiceband data transmission," *IEEE Journal on Selected Areas in Communications*, vol. 2, pp. 687–702, September 1984.

[131] G. D. Forney, "Multidimensional constellations. II: Voronoi constellations," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 941–958, August 1989.

[132] J.-E. Porath and T. Aulin, "Design of multidimensional signal constellations," *IEE Proceedings-Communications*, vol. 150, pp. 317–23, October 2003.

[133] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.

[134] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 451–460, March 1999.

[135] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Transactions on Communications*, vol. 49, pp. 1–4, January 2001.

[136] N. Sharma and C. B. Papadias, "Improved quasi-orthogonal codes through constellation rotation," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Orlando, FL, USA), March 2002.

[137] W. Su and X.-G. Xia, "Signal constellations for quasi-orthogonal space-time block codes with full diversity," *IEEE Transactions on Information Theory*, vol. 50, pp. 2331–2347, October 2004.

[138] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications-a comprehensive literature survey," *IEEE Communications Surveys & Tutorials*, vol. 11, pp. 87–105, June 2009.

[139] F. Oggier, C. Hollanti, and R. Vehkalahti, "An algebraic MIDO-MISO code construction," in *Proc. of International Conference on Signal Processing and Communications*, (Bangalore, India), July 2010.

[140] L. L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*, vol. 4. John Wiley & Sons, 2009.

[141] A. Dogandžic, "Chernoff bounds on pairwise error probabilities of space-time codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 1327–1336, May 2003.

[142] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.

[143] J. G. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-Hill, 5 ed., 2007.

[144] S. J. Wright and J. Nocedal, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 2006.

[145] R. J. Vanderbei and D. F. Shanno, "An interior-point algorithm for nonconvex nonlinear programming," *Computational Optimization and Applications*, vol. 13, pp. 231–252, April 1999.

[146] F. Kayhan and G. Montorsi, "Constellation design for memoryless phase noise channels," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 2874–2883, May 2014.

[147] J. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Transactions on Information Theory*, vol. 29, pp. 820–824, November 1983.

[148] G. Foschini, R. Gitlin, and S. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Transactions on Communications*, vol. 22, pp. 28–38, January 1974.

[149] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, pp. 632–647, September 1984.

[150] R. L. Graham and N. J. A. Sloane, "Penny-packing and two-dimensional codes," *Discrete & Computational Geometry*, vol. 5, pp. 1–11, January 1990.

[151] C. Thomas, M. Weidner, and S. Durrani, "Digital amplitude-phase keying with m-ary alphabets," *IEEE Transactions on Communications*, vol. 22, pp. 168–180, February 1974.

[152] M. Kifle and M. Vanderaar, "Bounds and simulation results of 32-ary and 64-ary quadrature amplitude modulation for broadband-ISDN via satellite," Tech. Rep. NASA-TM-106484, NASA Lewis Research Center, Cleveland, OH, USA, February 1994.

[153] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, pp. 951–963, April 2003.

[154] M. J. Borran, *Non-Coherent and Partially Coherent Space-Time Constellations*. PhD thesis, Rice University, Houston, TX, USA, 2003.

[155] M. O. Damen, A. Tewfik, and J.-C. Belfiore, "A construction of a space-time code based on number theory," *IEEE Transactions on Information Theory*, vol. 48, pp. 753–760, March 2002.

[156] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, October 1998.

[157] H. Jafarkhani and N. Seshadri, "Super-orthogonal space-time trellis codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 937–950, April 2003.

[158] S. Sezginer and H. Sari, "Full-rate full-diversity 2×2 space-time codes of reduced decoder complexity," *IEEE Communications Letters*, vol. 11, pp. 973–975, December 2007.

[159] K. Nikitopoulos, F. Mehran, and H. Jafarkhani, "Space-time super-modulation: Concept, design rules, and its application to joint medium access and rateless transmission," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 8275–8288, December 2017.

[160] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, pp. 3–18, January 1965.

[161] J. Wozencraft and R. Kennedy, "Modulation and demodulation for probabilistic coding," *IEEE Transactions on Information Theory*, vol. 12, pp. 291–297, July 1966.

[162] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 3233–3258, July 2006.

[163] P. R. Balogun, I. D. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Polar code design for irregular multidimensional constellations," *IEEE Access*, vol. 5, pp. 21941–21953, October 2017.

[164] B. R. Vojcic, L. B. Milstein, and R. L. Pickholtz, "Outage probability for the uplink of a DS-CDMA system operating over a LEO mobile satellite channel," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, (The Hague, Netherlands), September 1994.

[165] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, pp. 1804–1824, July 2002.

[166] V. Vakilian, H. Mehrpouyan, Y. Hua, and H. Jafarkhani, "High-rate space coding for reconfigurable 2×2 millimeter-wave MIMO systems," *arXiv:1505.06466*, May 2015.

[167] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 2228–2241, July 2008.

[168] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalised spatial modulation," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), November 2010.

[169] N. Ma, A. Wang, C. Han, and Y. Ji, "Adaptive joint mapping generalised spatial modulation," in *Proc. of IEEE International Conference on Communications in China*, (Beijing, China), August 2012.

[170] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. of IEEE International Conference on Neural Networks*, (Perth, WA, Australia), November 1995.

[171] M. Clerc, *Particle Swarm Optimization*. Newport Beach, CA, USA: Wiley-ISTE, 2010.

[172] A. Sezgin and E. A. Jorswieck, "Tight upper bound on the outage probability of QSTBC," *IEEE Communications Letters*, vol. 10, pp. 784–786, November 2006.

[173] N. Gresset, L. Brunel, and J. J. Boutros, "Space–time coding techniques with bit-interleaved coded modulations for MIMO block-fading channels," *IEEE Transactions on Information Theory*, vol. 54, pp. 2156–2178, May 2008.

[174] H. Holm and M.-S. Alouini, "Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution," *IEEE Transactions on Communications*, vol. 52, pp. 1367–1376, August 2004.

[175] M.-S. Alouini, A. Abdi, and M. Kaveh, "Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels," *IEEE Transactions on Vehicular Technology*, vol. 50, pp. 1471–1480, November 2001.

[176] Y. Feng, M. Wen, J. Zhang, F. Ji, and G. Ning, "Sum of arbitrarily correlated gamma random variables with unequal parameters and its application in wireless communications," in *Proc. of IEEE International Conference on Computing, Networking, and Communication*, (Kauai, HI, USA), February 2016.

[177] C. O. Archer, "Some properties of Rayleigh distribution random variables and of their sums and products," Tech. Rep. TM-7-15, Naval Missile Center, Point Mugu, CA, USA, April 1967.

[178] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Transactions on Communications*, vol. 48, pp. 948–959, June 2000.

[179] M. Yazdani and A. H. Banihashemi, "On construction of rate-compatible low-density parity-check codes," in *Proc. of IEEE International Conference on Communications*, (Paris, France), June 2004.

[180] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Transactions on Communications*, vol. 52, pp. 1311–1321, August 2004.

[181] K. Niu, K. Chen, and J.-R. Lin, "Beyond turbo codes: rate-compatible punctured polar codes," in *Proc. of IEEE International Conference on Communications*, (Budapest, Hungary), June 2013.